ESCAP MEETING NO. 37 - 02/13/01

AGENDA

Kathleen P Porter
02/13/2001 09:17 AM

To: Angela Frazier/DMD/HQ/BOC@BOC, Annette M Quinlan/DMD/HQ/BOC@BOC, Barbara E Hotchkiss/DSD/HQ/BOC@BOC, Betty Ann Saucier/DIR/HQ/BOC@BOC, Carnelle E Sligh/PRED/HQ/BOC@BOC, Carol M Van Horn/DMD/HQ/BOC@BOC, Carolee Bush/DMD/HQ/BOC@BOC, Cynthia Z F Clark/DIR/HQ/BOC@BOC, Deborah A Fenstermaker/DSSD/HQ/BOC@BOC, Donna L Kostanich/DSSD/HQ/BOC@BOC, Hazel V Beaton/SRD/HQ/BOC@BOC, Howard R Hogan/DSSD/HQ/BOC@BOC, John F Long/POP/HQ/BOC@BOC, John H Thompson/DMD/HQ/BOC@BOC, Kathleen M Styles/DMD/HQ/BOC@BOC, Linda A Hiner/DSSD/HQ/BOC@BOC, Lois M Kline/POP/HQ/BOC@BOC, Margaret A Applekamp/DIR/HQ/BOC@BOC, Maria E Urrutia/DMD/HQ/BOC@BOC, Marvin D Raines/DIR/HQ/BOC@BOC, Mary A Cochran/DIR/HQ/BOC@BOC, Mary E Williams/DIR/HQ/BOC@BOC, Nancy A Potok/DIR/HQ/BOC@BOC, Nancy M Gordon/DSD/HQ/BOC@BOC, Nicholas I Birnbaum/DMD/HQ/BOC@BOC, Patricia E Curran/DIR/HQ/BOC@BOC, Paula J Schneider/DIR/HQ/BOC@BOC, Phyllis A Bonnette/DIR/HQ/BOC@BOC, Preston J Waite/DMD/HQ/BOC@BOC, Rajendra P Singh/DSSD/HQ/BOC@BOC, Robert E Fay III/DIR/HQ/BOC@BOC, Ruth Ann Killion/PRED/HQ/BOC@BOC, Sarah E Brady/DMD/HQ/BOC@BOC, Sue A Kent/DMD/HQ/BOC@BOC, Tommy Wright/SRD/HQ/BOC@BOC, Vanessa M Leuthold/DMD/HQ/BOC@BOC, William G Barron Jr/DIR/HQ/BOC@BOC

cc: Michael D Starsinic/DSSD/HQ/BOC@BOC, Mary Helen Mulry/DSSD/HQ/BOC@BOC

Subject: Agenda for 2/13

The agenda for the February 13 ESCAP Meeting scheduled from 10:30-12 in Rm. 2412/3 is as follows:

1. Variance Estimates by Size of Geographic Area - Mike Starsinic

2. Total Error Model Results - Mary Mulry

ESCAP MEETING NO. 37 - 02/13/01

HANDOUTS

Materials attached to these minutes were draft and preliminary material to inform the ESCAP Committee. The data and analysis contained in these documents are subject to revision and are not final. These materials report the results of research and analysis undertaken by Census Bureau staff. They have undergone a more limited review than official Census Bureau publications. Research results and conclusions expressed are those of the authors and do not necessarily indicate concurrence by the Census Bureau.

OVERVIEW OF RESULTS FOR TOTAL ERROR MODEL

Mary H. Mulry
February 13, 2001

Purpose of total error analysis

- We review the net effect of the sampling and nonsampling components of error in the A.C.E., including varying assumptions about correlation bias.
- We review confidence intervals for the A.C.E. estimates based on bias and variance estimates from the total error analysis prior to using the methodology in the loss function analysis comparing the original enumeration and the A.C.E. estimates.

Estimation Strategy

- First, we estimate component errors and their variances for groups of A.C.E. poststrata called evaluation poststrata, some with 1990 data.
- Then we derive estimates of component errors for each A.C.E. postratum based on the component errors for its evaluation poststrata.
- We use simulation methodology to assess the net effect of all the component errors combined and for use in the loss function analysis.

Results

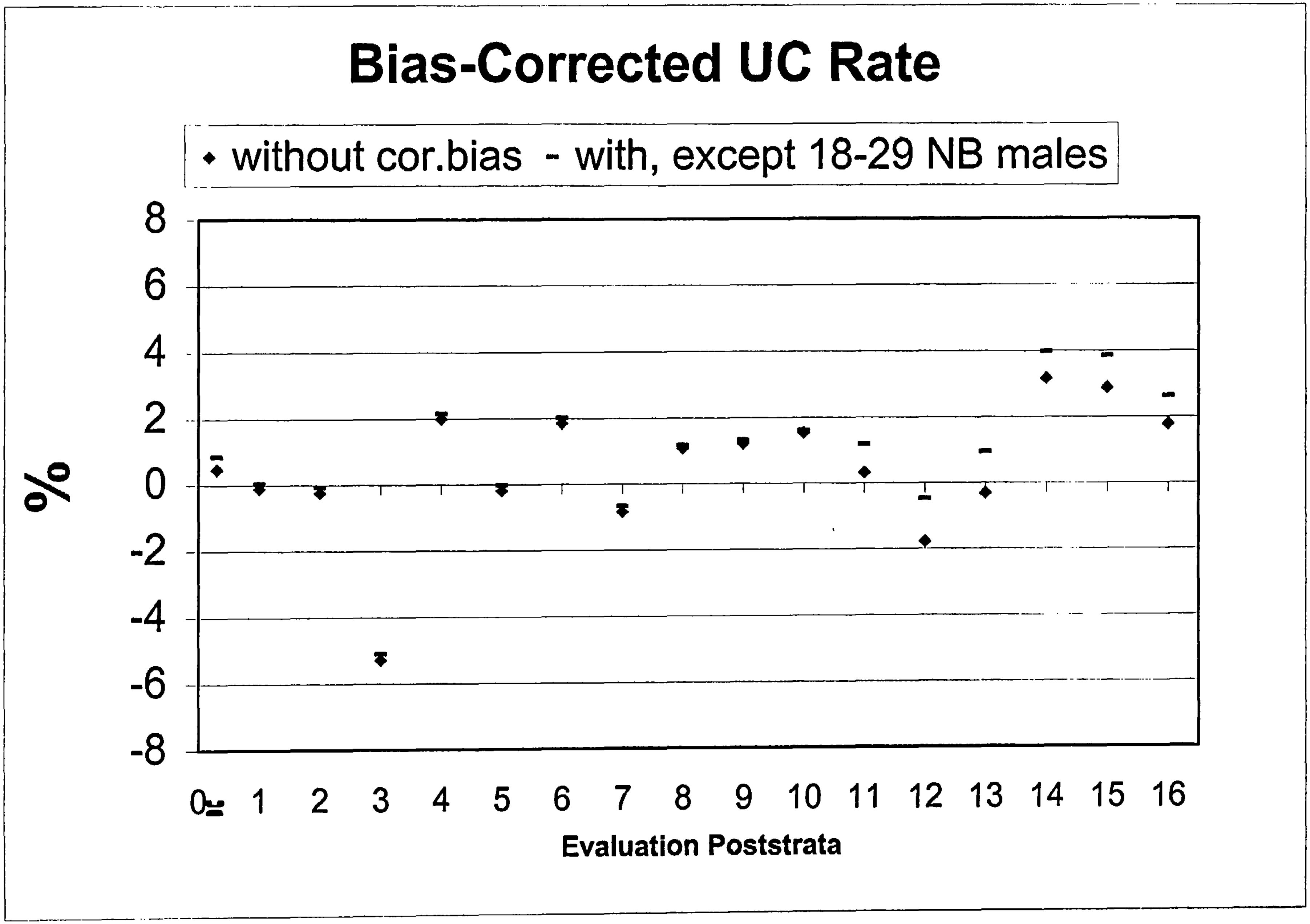
- Assumptions about correlation bias affect the number of confidence intervals that cross zero.
- Even so, the estimation of the component errors indicates that undercount estimates based on the DSE is not the result of noise, but a real phenomenon.

Components of Error

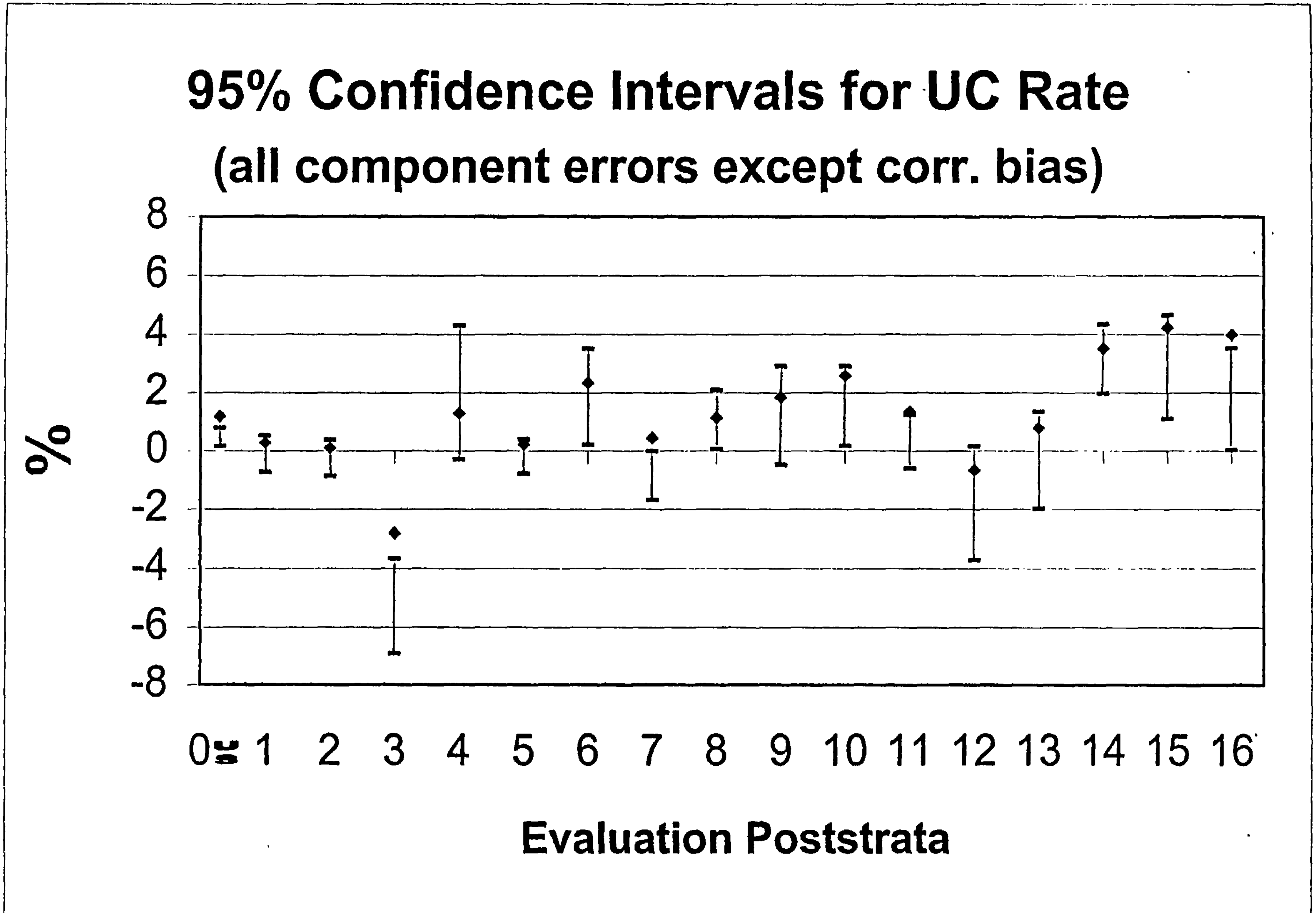
Error Components	Measurement in 1990	Measurement in 2000
P-sample matching error	1990 Matching Error Study	1990 Matching Error Study with adjustments for 2000
P-sample data collection error	1990 Evaluation Followup	1990 Evaluation Followup with adjustments for 2000
P-sample fabrication	1990 Evaluation Followup	1990 Evaluation Followup with adjustments for 2000
E-sample data collection error	1990 Evaluation Followup	1990 Evaluation Followup with adjustments for 2000
E-sample processing error	1990 Matching Error Study	1990 Matching Error Study with adjustments for 2000
Correlation bias	1990 Demographic Analysis	2000 Demographic Analysis
Ratio estimator bias	1990 PES	2000 A.C.E.
Sampling error	1990 PES	2000 A.C.E.
Imputation error	1990 Reasonable Alternatives Imputation Study	1990 Reasonable Alternatives with adjustments for 2000
Excluded Census Data Error	1990 Excluded Data Study	Not available
Contamination of P sample by enumeration or vice versa	Shown to be negligible	Not available in time for analysis for decision
Misclassification error of records into poststrata from inconsistent reporting	Not measured	Not available in time for analysis for decision
Synthetic error	Artificial population analysis and not integrated in total error model	Under development but will not be integrated in total error model

Assumptions and Limitations:

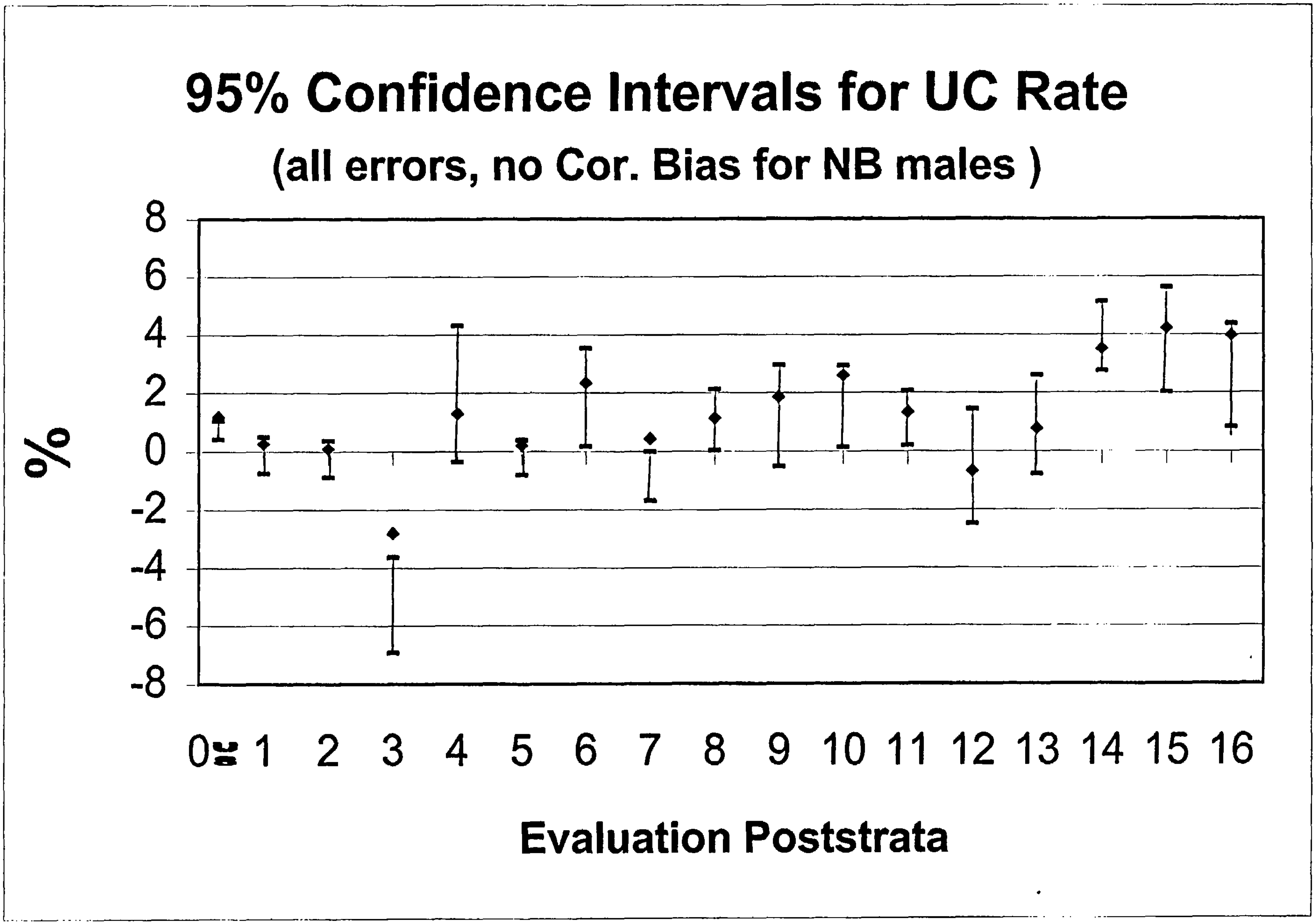
- The assumption for nonsampling error components from field and processing operations is that the errors measured in the 1990 PES scaled for the 2000 population reasonably reflect the errors for the 2000 A.C.E.
- The mapping of the 1990 PES poststrata to the 2000 A.C.E. poststrata uses characteristics from the 1990 census.
- The sex ratios from demographic analysis are reliable enough to use in estimating correlation bias in the DSE, and there is no correlation bias present for females.
- The assumptions about correlation bias considered:
    - No correlation bias is present in the DSE.
    - Correlation bias is present for Black males but not for Non-black males.
      Correlation bias is present for all males except Non-black males 18 to 29 years of age.
    - Correlation bias is present in the DSE as measured by the sex ratios from demographic analysis, including a 2% overcount of 18 to 29 year-old Non-black males by the DSE.

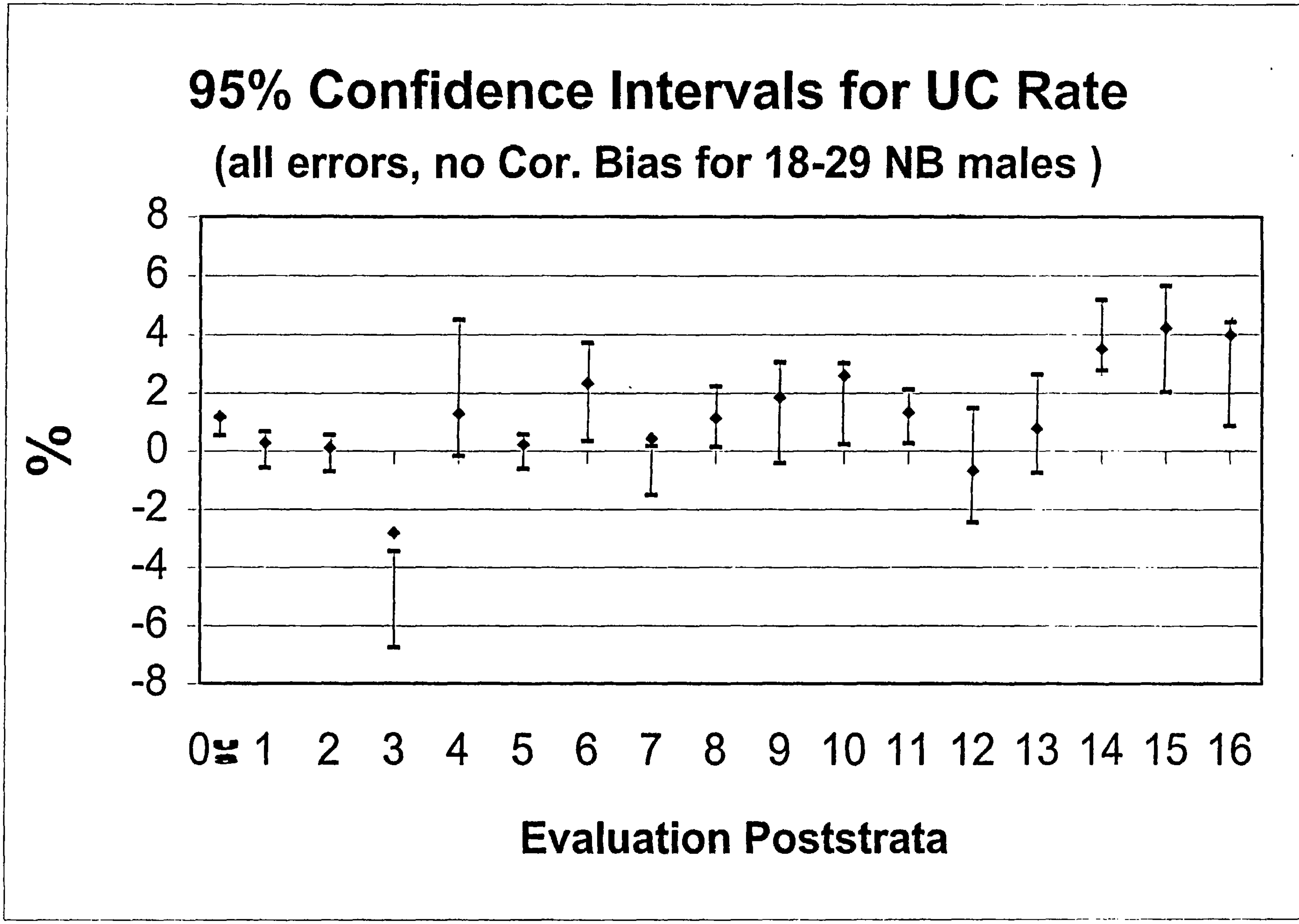


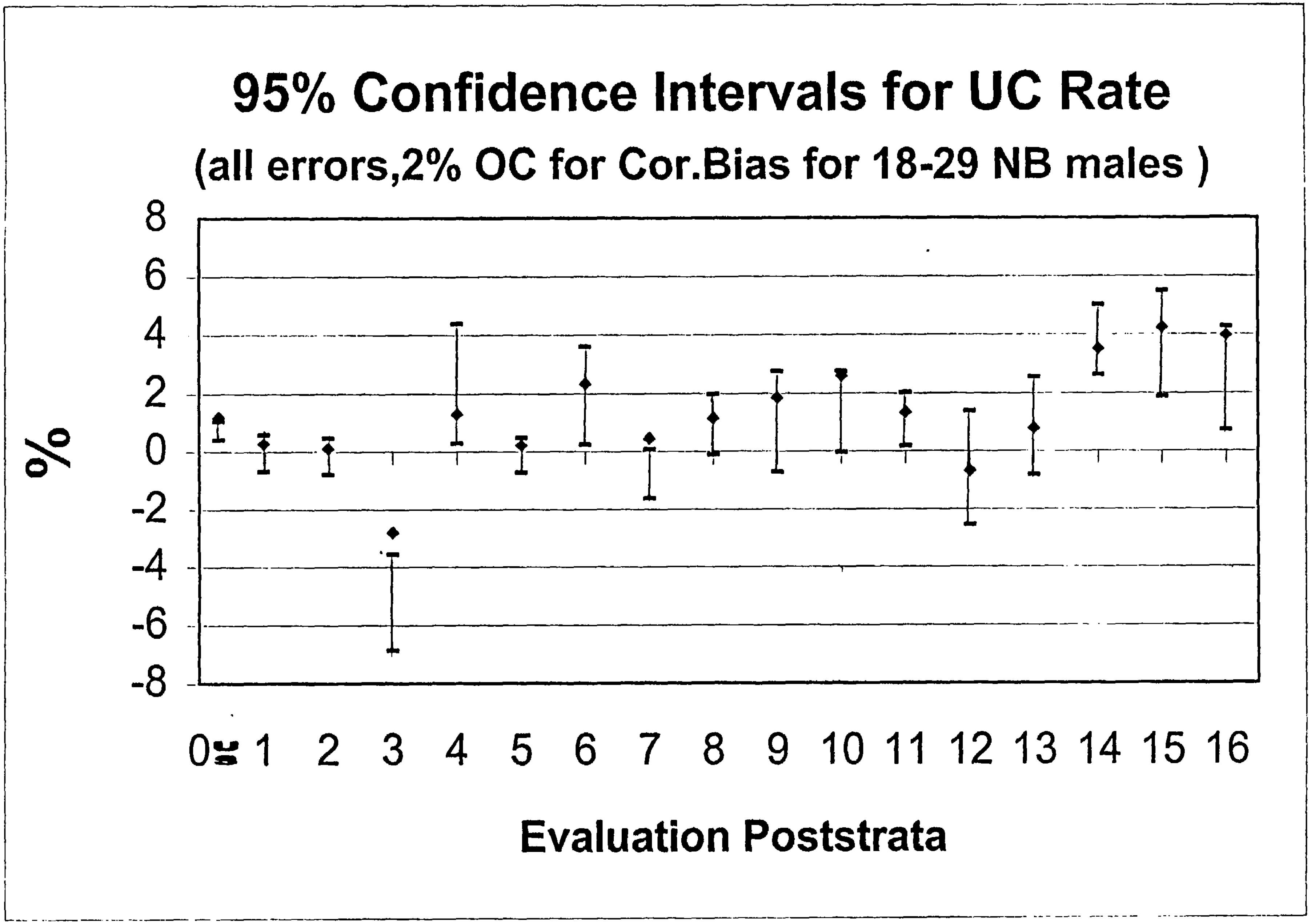
Bias-Corrected UC Rate
◆ without cor.bias - with, except 18-29 NB males
8
6
4
2
0
-2
-4
-6
-8
%
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
Evaluation Poststrata



95% Confidence Intervals for UC Rate
(all component errors except corr. bias)
%
8
6
4
2
0
-2
-4
-6
-8
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
Evaluation Poststrata


95% Confidence Intervals for UC Rate
(all errors, no Cor. Bias for NB males)
8
6
4
2
0
-2
-4
-6
-8
%
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
Evaluation Poststrata


95% Confidence Intervals for UC Rate
(all errors, no Cor. Bias for 18-29 NB males)
8
6
4
2
0
-2
-4
-6
-8
%
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
Evaluation Poststrata


95% Confidence Intervals for UC Rate
(all errors,2% OC for Cor.Bias for 18-29 NB males)
8
6
4
2
0
-2
-4
-6
-8
%
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
Evaluation Poststrata

Confidence intervals without correlation bias

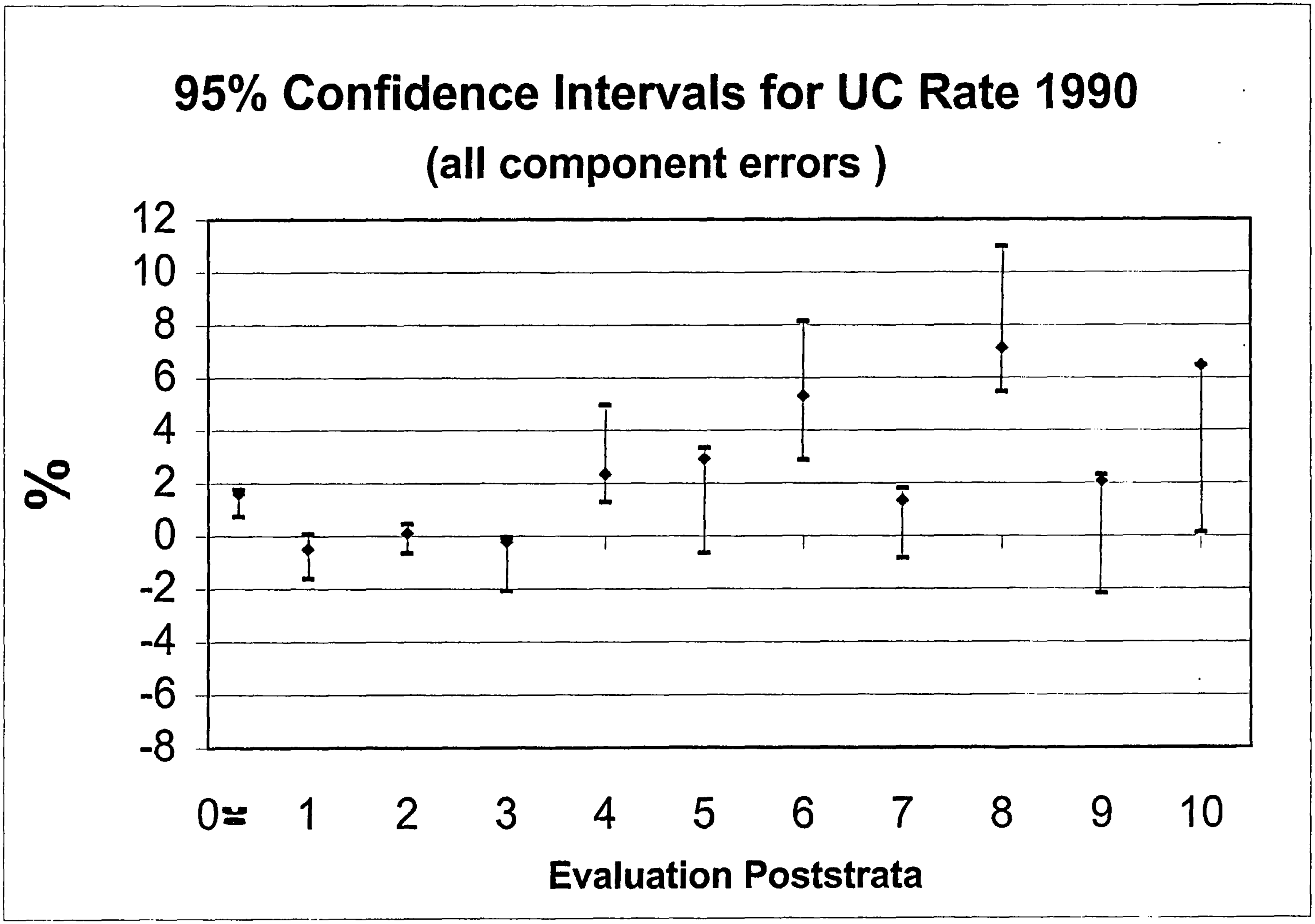
evaluation poststrata	prod UC	95% conf interval
US	1.1788	(0.1478 , 0.7818)
1.N-min/own/lrg&med MSA/MO-MB-hi/NE/MW	0.2695	(-0.7400 , 0.5014)
2.N-min/own/lrg&med MSA/MO-MB-hi/S/W	0.0947	(-0.8685 , 0.3665)
3.N-min/own/lrg&med MSA/MO-MB-lo/NE/MW	-2.8191	(-6.9126 , -3.6736)
4.N-min/own/lrg&med MSA/MO-MB-lo/S/W	1.2840	(-0.3111 , 4.2940)
5.N-min/own/small MSA/MO-MB-hi	0.2127	(-0.7920 , 0.3828)
6.N-min/own/small MSA/MO-MB-lo	2.3302	(0.2014 , 3.5071)
7.N-min/own/All other TEAs	0.4232	(-1.6867 , -0.0198)
8.N-min/n-own/lrg&med MSA/MO-MB-hi	1.1290	(0.0483 , 2.0913)
9.N-min/n-own/lrg&med MSA/MO-MB-lo	1.8404	(-0.4858 , 2.9104)
10.N-min/n-own/small MSA/MO-MB&other TEA	2.5867	(0.1562 , 2.1390)
11.Min/own/lrg&med MSA/MO-MB-hi	1.3307	(-0.9075 , 1.2351)
12.Min/own/lrg&med MSA/MO-MB-lo	-0.6778	(-3.7351 , 0.1517)
13.Min/own/All other TEAs	0.7719	(-1.9900 , 1.3370)
14.Min/n-own/lrg&med MSA/MO-MB-hi	3.5018	(1.9694 , 4.3420)
15.Min/n-own/lrg&med MSA/MO-MB-lo	4.2140	(1.0884 , 4.6401)
16.Min/n-own/All other TEAs	3.9699	(0.0172 , 3.5212)

Confidence intervals with correlation bias

eval ps	prod UC	95% conf interval no error for 18-29 NB males	95% conf interval 2% overct for cor.bias 18-29 NB M
US	1.1788	(0.537 , 1.1763)	(0.4013 , 1.0407)
1.N-min/own/lrg&med MSA/MO-MB-hi/NE/MW	0.2695	(-0.583 , 0.666)	(-0.6798 , 0.5702)
2.N-min/own/lrg&med MSA/MO-MB-hi/S/W	0.0947	(-0.701 , 0.5477)	(-0.7882 , 0.4606)
3.N-min/own/lrg&med MSA/MO-MB-lo/NE/MW	-2.8191	(-6.752 , -3.451)	(-6.8725 , -3.57)
4.N-min/own/lrg&med MSA/MO-MB-lo/S/W	1.284	(-0.175 , 4.4915)	(0.2808 , 4.3865)
5.N-min/own/small MSA/MO-MB-hi	0.2127	(-0.625 , 0.5681)	(-0.7215 , 0.4712)
6.N-min/own/small MSA/MO-MB-lo	2.3302	(0.3357 , 3.7086)	(0.2283 , 3.6015)
7.N-min/own/All other TEAs	0.4232	(-1.516 , 0.171)	(-1.6214 , 0.0664)
8.N-min/n-own/lrg&med MSA/MO-MB-hi	1.129	(0.1302 , 2.2182)	(-0.1155 , 1.9734)
9.N-min/n-own/lrg&med MSA/MO-MB-lo	1.8404	(-0.417 , 3.0522)	(-0.715 , 2.7578)
10.N-min/n-own/small MSA/MO-MB&other TEA	2.5867	(0.2257 , 3.0173)	(-0.0438 , 2.752)
11.Min/own/lrg&med MSA/MO-MB-hi	1.3307	(0.2649 , 2.1163)	(0.1679 , 2.0185)
12.Min/own/lrg&med MSA/MO-MB-lo	-0.6778	(-2.452 , 1.4778)	(-2.5483 , 1.3802)
13.Min/own/All other TEAs	0.7719	(-0.742 , 2.6327)	(-0.819 , 2.552)
14.Min/n-own/lrg&med MSA/MO-MB-hi	3.5018	(2.7808 , 5.1699)	(2.6112 , 5.0008)
15.Min/n-own/lrg&med MSA/MO-MB-lo	4.214	(2.0345 , 5.6516)	(1.8705 , 5.4889)
16.Min/n-own/All other TEAs	3.9699	(0.8592 , 4.4078)	(0.7097 , 4.2594)

16 Evaluation Poststrata

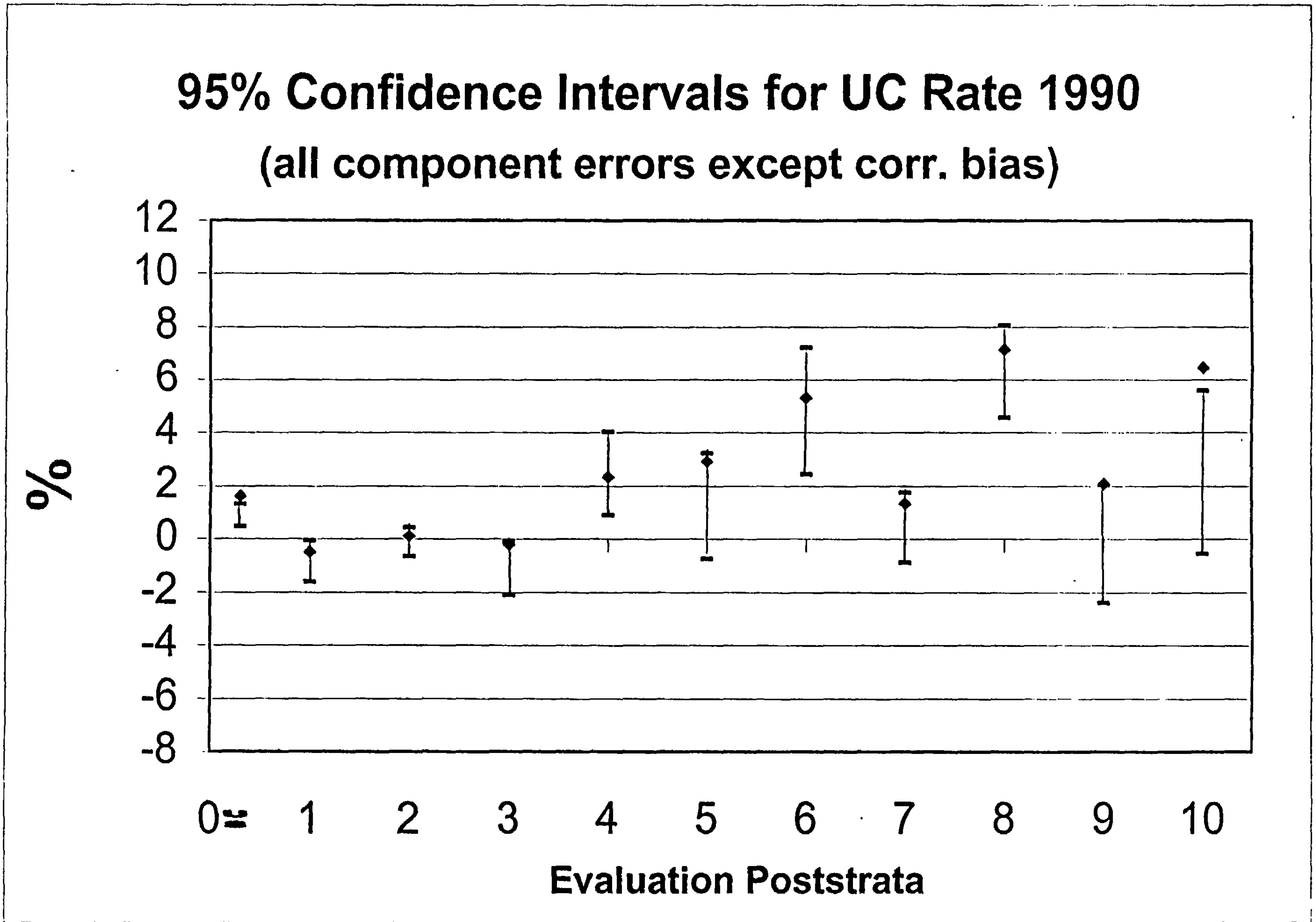
		No. in MVF P-sample (1990)	PS Groups (2000)
1. Non-minority/owner/large and Medium MSA MO-MB NE/MW	high RR	4,960	1-4
2. Non-minority/owner/large and Medium MSA MO-MB S/W	high RR	7,702	9-12
3. Non-minority/owner/large and Medium MSA MO-MB NE/MW	low RR	3,031	5,6,13,14
4. Non-minority/owner/large and Medium MSA MO-MB S/W	low RR	2,936	7,8,15,16
5. Non-minority/owner/Small MSA and Non-MSA MO-MB	high RR	5,560	17-20
6. Non-minority/owner/ Small MSA and Non-MSA MO-MB	low RR	2,095	21-24
7. Non-minority/Owner/All Other TEAs		7,355	25-32
8. Non-minority/ Non-Owner/Large or Medium MSA MO-MB	high RR	4,963	33, 35
9.Non-minority/ Non-Owner/Large or Medium MSA MO-MB	low RR	3,197	34, 36
10. Non-minority/non-owner/Small MSA & Non-MSA MO-MB All other TEA		5,291	37-40
11. Minority/owner/large and Medium MSA MO-MB	high RR	8,841	41, 49, 57,59
12. Minority/owner/large and Medium MSA MO-MB	low RR	5,628	42, 50
13. Minority/Owner/All Other TEAs		3,877	43, 44, 51, 52
14. Minority/ Non-Owner/Large or Medium MSA MO-MB	high RR	10,809	45, 53, 58,60
15. Minority/ Non-Owner/Large or Medium MSA MO-MB	low RR	6,421	46, 54
16. Minority/Non-Owner/All Other TEAs		3,797	47, 48, 55, 56, 61-64
Total		86,463	



95% Confidence Intervals for UC Rate 1990
(all component errors)
%
12
10
8
6
4
2
0
-2
-4
-6
-8
0
1
2
3
4
5
6
7
8
9
10
Evaluation Poststrata

Table 3.1. The 10 Evaluation Poststrata – 1990

1	Non-Hispanic White and Other, Owner in Urban Areas 250k+
2	Non-Hispanic White and Other, Owner in Other Urban Areas
3	Non-Hispanic White and Other, Owner in Non-Urban Areas
4	Non-Hispanic White and Other, Non-Owner in Urban Areas 250k+
5	Non-Hispanic White and Other, Non-Owner in Other Urban Areas
6	Non-Hispanic white and Other, Non-Owner in Non-Urban Areas
7	Black, Non-Black Hispanic, Asian and Pacific Islander, Owner in Urban Areas 250k+
8	Black, Non-Black Hispanic, Asian and Pacific Islander, Non-Owner in Urban Areas 250k+
9	Black, Non-Black Hispanic, Asian and Pacific Islander Owner in Other Urban & Non-Urban Areas
10	Black, Non-Black Hispanic, Asian and Pacific Islander Non-Owner in Other Urban & Non-Urban Areas


95% Confidence Intervals for UC Rate 1990
(all component errors except corr. bias)
12
10
8
6
4
2
0
-2
-4
-6
-8
%
0
1
2
3
4
5
6
7
8
9
10
Evaluation Poststrata

Materials attached to these minutes were draft and preliminary material to inform the ESCAP Committee. The data and analysis contained in these documents are subject to revision and are not final. These materials report the results of research and analysis undertaken by Census Bureau staff. They have undergone a more limited review than official Census Bureau publications. Research results and conclusions expressed are those of the authors and do not necessarily indicate concurrence by the Census Bureau.

Table 1: US Summary of Distribution of CVs for Population Estimates by Geographical Area for 1990 PES and 2000 A.C.E

Area	Source (1)	Number (2)	Mean Size (3)	Mean CV (4)	Margin of Error* (5)	Disitribution of CVs: Minimum (6)	Q1 (7)	Median (8)	Q3 (9)	Maximum (10)
State **	A.C.E.	51	5,582,035	0.310%	28,506	0.159%	0.220%	0.240%	0.378%	0.804%
	PES	51	4,955,153	0.449%	36,623	0.322%	0.369%	0.406%	0.496%	0.933%
Congressional Disitricts ***	A.C.E.	435	653,103	0.330%	3,546	0.156%	0.250%	0.297%	0.375%	0.948%
	PES	435	579,567	0.557%	5,309	0.299%	0.420%	0.499%	0.628%	2.007%
Places > 100,000 ****	A.C.E.	245	315,037	0.343%	1,776	0.213%	0.283%	0.314%	0.361%	1.435%
	PES	195	335,637	0.673%	3,718	0.363%	0.536%	0.629%	0.747%	1.702%
Counties > 100,000 ****	A.C.E.	524	409,345	0.368%	2,481	0.201%	0.274%	0.310%	0.405%	1.498%
	PES	458	400,593	0.534%	3,519	0.285%	0.432%	0.510%	0.591%	1.483%

* - Margin of Error is calculated as 1.645 * standard error of the population estimate.

** - “State” includes all 50 states and the District of Columbia, but does not include Puerto Rico.

*** - 103rd Congressional Districts for the PES; 106th Congressional Districts for the A.C.E. Does not include the District of Columbia or Puerto Rico.

**** - Counties and places with census counts of more than 100,000 in the respective censuses, 2000 for A.C.E. and 1990 for PES.

Materials attached to these minutes were draft and preliminary material to inform the ESCAP Committee. The data and analysis contained in these documents are subject to revision and are not final. These materials report the results of research and analysis undertaken by Census Bureau staff. They have undergone a more limited review than official Census Bureau publications. Research results and conclusions expressed are those of the authors and do not necessarily indicate concurrence by the Census Bureau.

Accuracy & Coverage Evaluation: Variance Estimates by Size of Geographic Area
Summary of Results

Presented to ESCAP, February 13, 2001
Michael D. Starsinic, DSSD

Overall:

As expected, the coefficients of variation (CV's) were lower than the corresponding 1990 CV's for all four geographic areas we investigated. This was expected because:

- The housing unit sample size for the A.C.E. was almost double that of the PES (300,913 versus approximately 165,000).
- Better measures of population size were available during sample selection of clusters.
- Reduced variability of sampling weights.

All three of these improvements should lead to smaller sampling variances.

States: (Table 1, Graph 1)

- Median CV decreased by about 40%, from 0.406% to 0.240%.

Congressional Districts: (Table 1, Graph 2)

- Median CV dropped from 0.499% to 0.297%, about a 40% decrease.

Places > 100,000: (Table 1, Graph 3)

- Cutoff of 100,000 based on 1990 and 2000 population counts.

- Median CV dropped from 0.629% to 0.314%, about a 50% decrease.

Counties > 100,000: (Table 1, Graph 4)

- Cutoff of 100,000 based on 1990 and 2000 population counts.

- Median CV dropped from 0.510% to 0.310%, about a 40% decrease.

Ratio of Simulated to Production CVs (Graph 5)

- Graph 6 shows the ratio of simulated state-level CVs from loss function analysis to actual values. The ratio of the CVs is in the interval 0.97 to 1.03 for 47 states. In 1990, 42 states were *outside* this interval.

Appendix: Variance Estimation Methodology

The A.C.E. survey was a multi-phase sample, which increased the difficulties of estimating the sampling variance. Multi-phase sampling differs from multi-stage in the following way: in a multi-stage design, the information needed to draw all stages of the sample is known before the sampling begins; in a multi-phase design, the information needed to draw the n^{th} phase of the sample is unobtainable until the $n\text{-}1^{st}$ phase of the sample is completed. A methodology based in part on the Rao-Shao jackknife variance estimator (Rao & Shao 1992) takes into account the multi-phase nature of the A.C.E. The estimation of the variance due to the A.C.E. attempts to capture these components of the variance (the relative contribution to the sampling error from the components is not considered in this analysis):

- Sampling variance due to the initial Listing sample.

- Sampling variance due to the A.C.E. Reduction and Small Block Subsampling.

- Sampling variance due to the Targeted Extended Search (TES) sample.

- Variance due to the imputation of correct enumeration, match, and residence probabilities for unresolved cases.

This estimate of variance is only intended to include the error from the above four components, and is not intended to quantify nonsampling errors, other than the probability imputation error. Specific components of error which are *not* incorporated into the variance estimates are the synthetic error, the error due to weight trimming, and the error due to large block subsampling.

This new methodology directly estimates variances only for the final collapsed post-strata. We compute all other variances using a variance-covariance matrix for the post-stratum coverage correction factors (CCFs), which is the output of the variance estimation process (along the post-stratum variances). The estimated ("synthetic") variance of any population estimate can be computed using this matrix and the unadjusted census counts, broken down by post-stratum and excluding persons out-of-scope of the A.C.E. (For more information see Kim et al (2000) and Starsinic & Kim (2000).)

$$
\begin{aligned}
\hat{X}_s &= \text{Synthetic household population estimate for geographic area s} \\
&= \sum_{\text{post-strata } h} \hat{X}_{sh} \\
&= \sum_{h=1}^{416} C_{sh} \times CCF_h, \ \text{where } C_{sh} = \text{Census count of post-stratum h in geographic area s}
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{Var}(\hat{X}_s) &= \text{synthetic variance for synthetic household population estimate } \hat{X}_s \\
&= \mathrm{Var}\Big(\sum_{h=1}^{416} \hat{X}_{sh}\Big) \\
&= \sum_{h=1}^{416} \sum_{h'=1}^{416} \mathrm{Cov}(\hat{X}_{sh}, \hat{X}_{sh'}) \\
&= \sum_{h=1}^{416} \sum_{h'=1}^{416} \mathrm{Cov}(C_{sh} \times CCF_h, C_{sh'} \times CCF_{h'}) \\
&= \sum_{h=1}^{416} \sum_{h'=1}^{416} C_{sh} \times C_{sh'} \times \mathrm{Cov}(CCF_h, CCF_{h'})
\end{aligned}
$$

For any desired population estimate, geographic or otherwise:

$$
\text{Synthetic total population estimate} = \text{Synthetic household population estimate } (\hat{X}) + \text{“Residual” count}
$$

where the “Residual” count are persons out-of-scope of the A.C.E. sample. These include institutionalized and non-institutionalized group quarters persons; persons counted in Service Based Enumeration (SBE), and those estimated by the SBE’s multiplicity estimator; and persons enumerated in the Remote Alaska operation. (Variance due to the SBE’s multiplicity estimation is not accounted for in the A.C.E. variance estimates.)

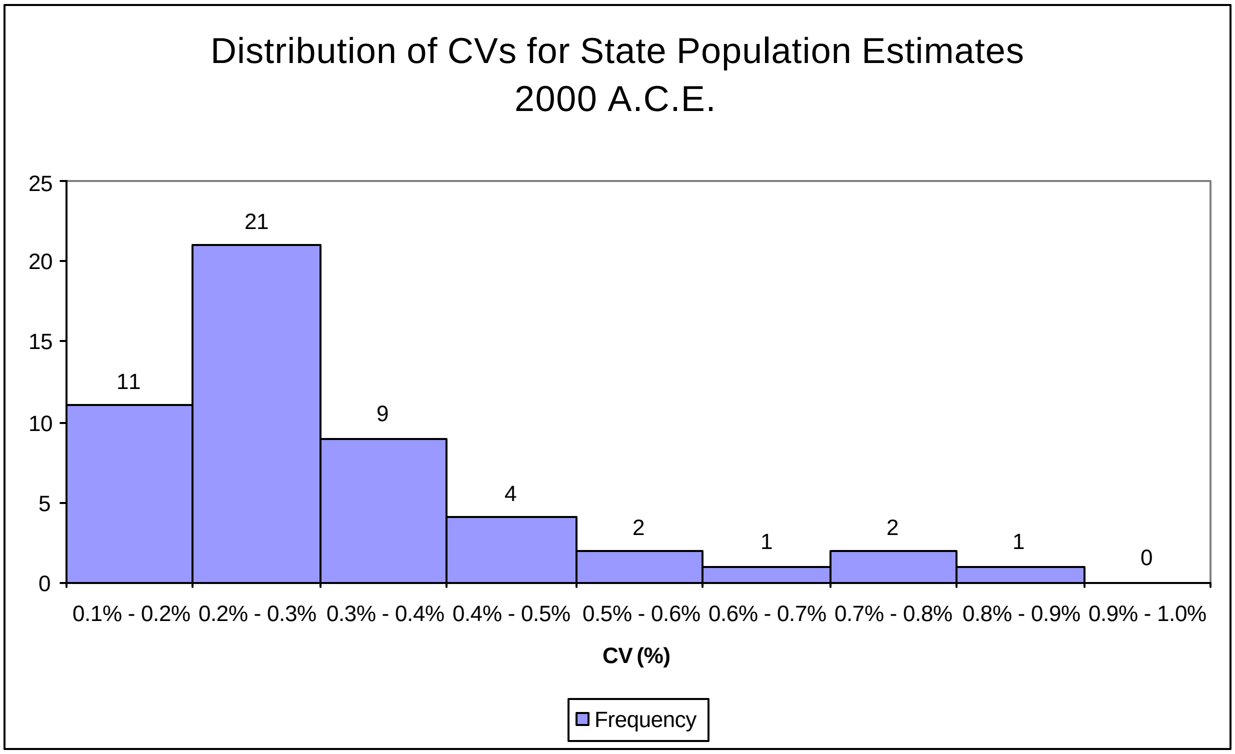
The coefficient of variation (CV) is computed as:

$$
CV = \frac{\sqrt{\mathrm{Var}(\text{Synthetic total population estimate})}}{\text{Synthetic total population estimate}}
$$

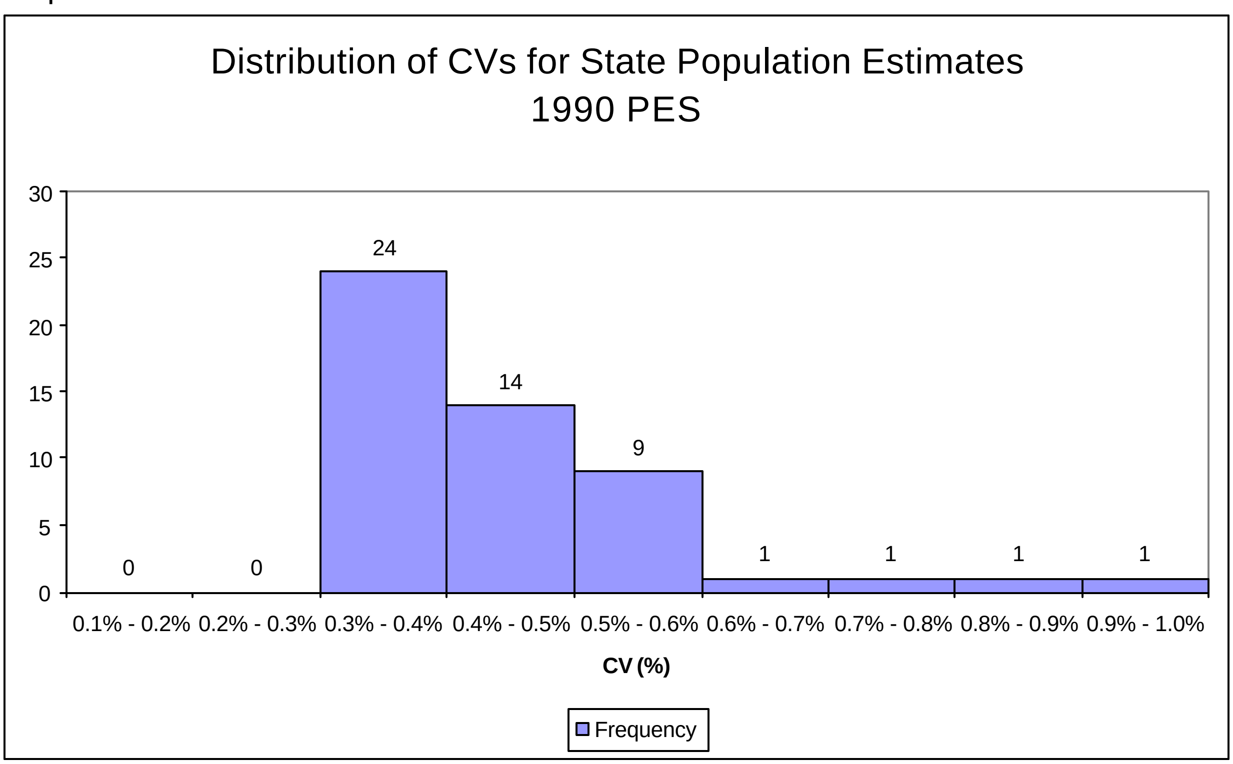
Since the Residual population is excluded from the A.C.E. sample, it adds no sampling variance, and the variance of the synthetic estimate is the same as the variance of the corresponding A.C.E estimate described above.

Materials attached to these minutes were draft and preliminary material to inform the ESCAP Committee. The data and analysis contained in these documents are subject to revision and are not final. These materials report the results of research and analysis undertaken by Census Bureau staff. They have undergone a more limited review than official Census Bureau publications. Research results and conclusions expressed are those of the authors and do not necessarily indicate concurrence by the Census Bureau.

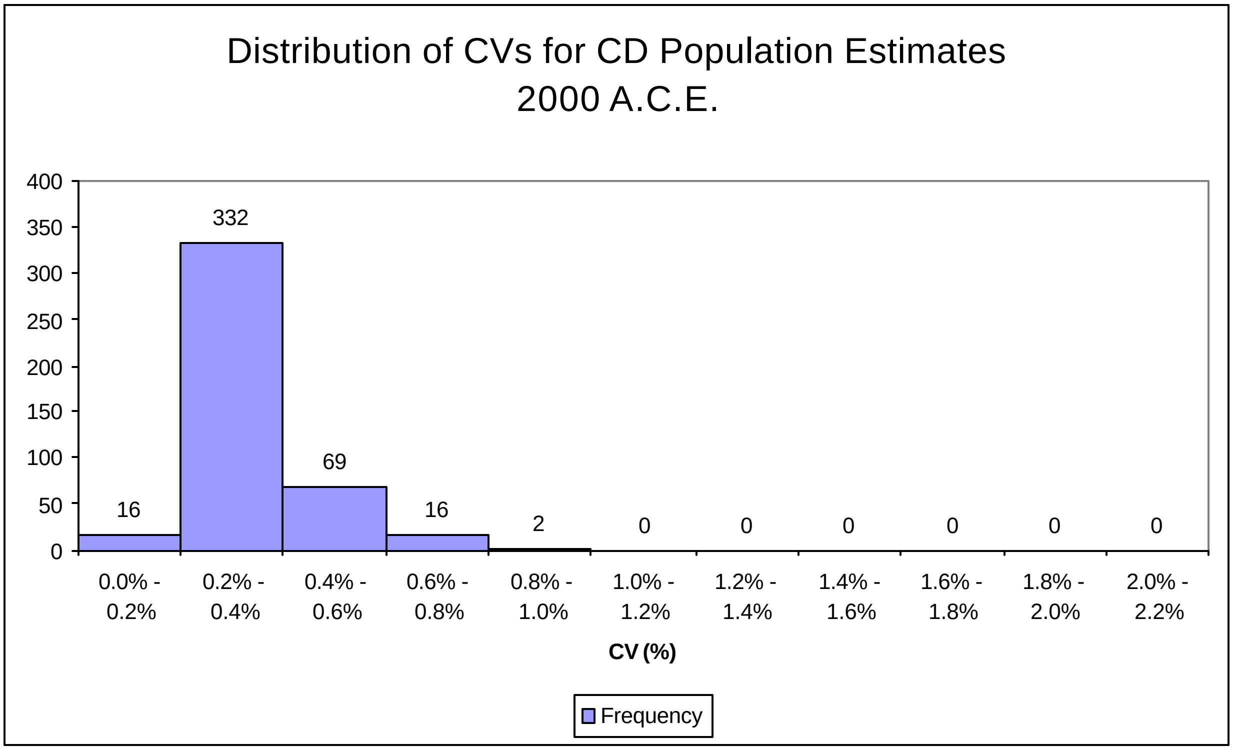
Graph 1a:



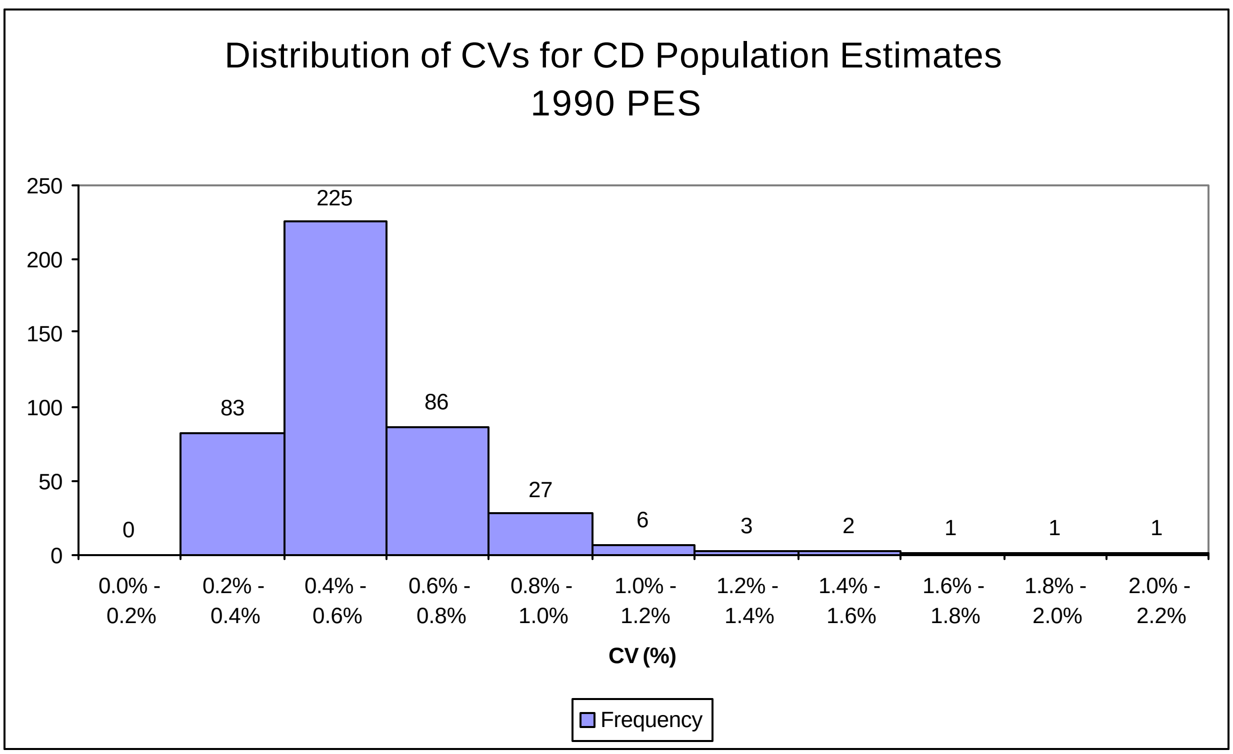
Graph 1b:



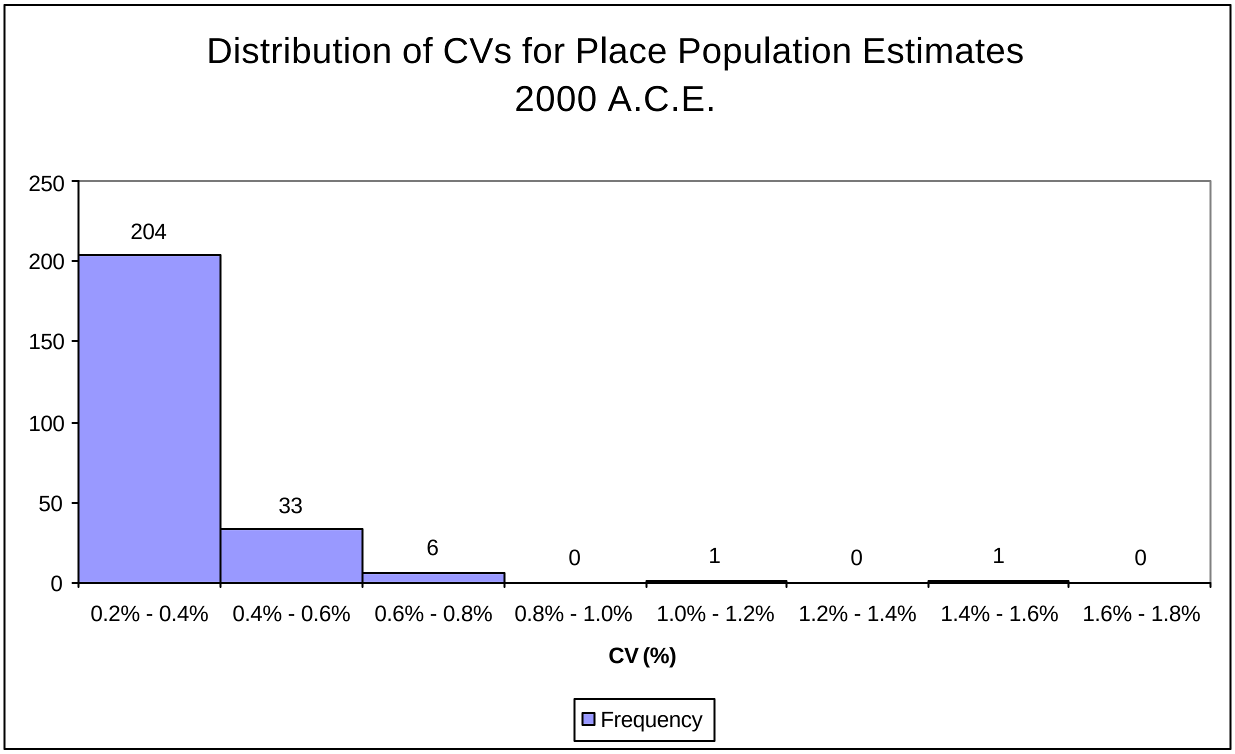
Graph 2a:



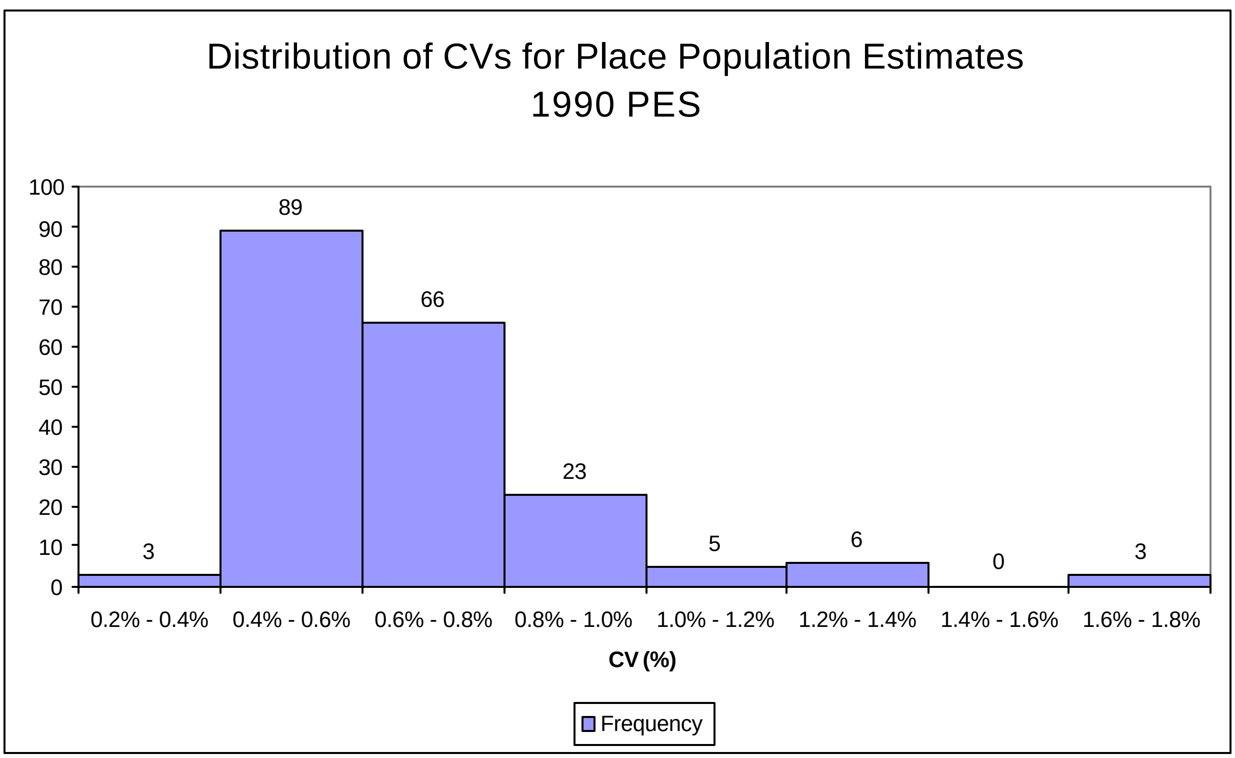
Graph 2b:



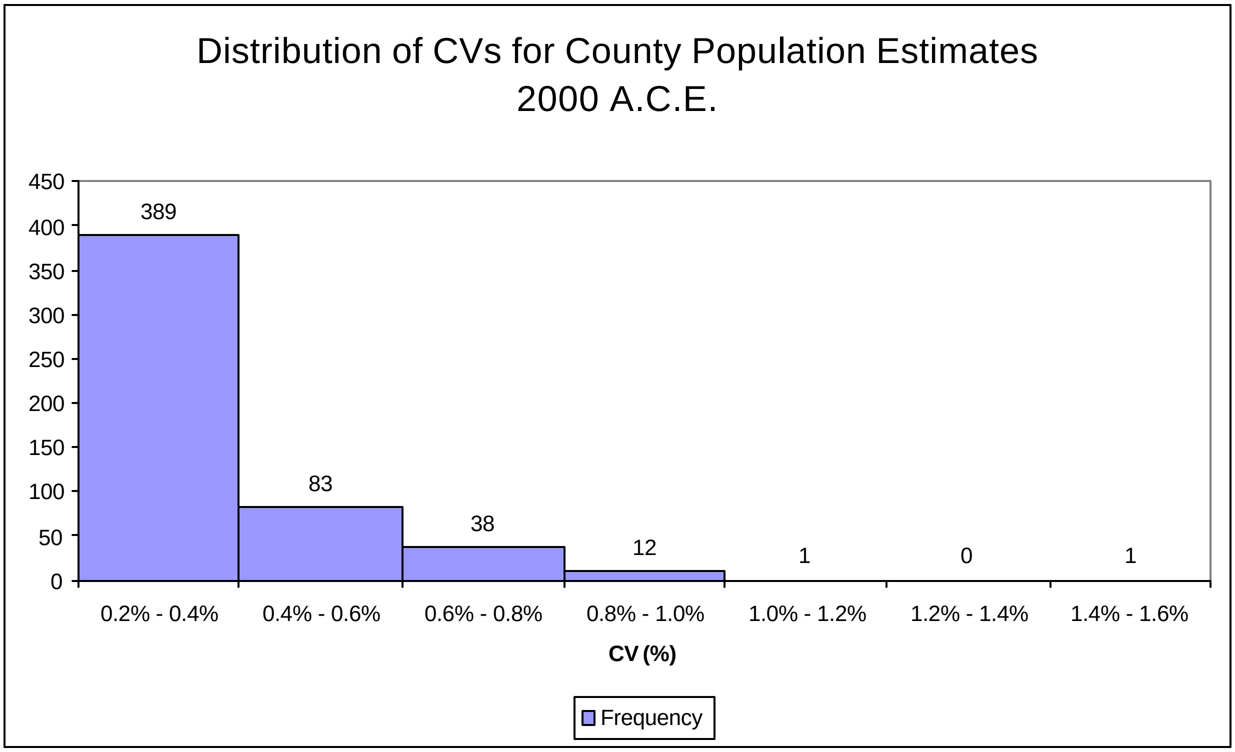
Graph 3a:



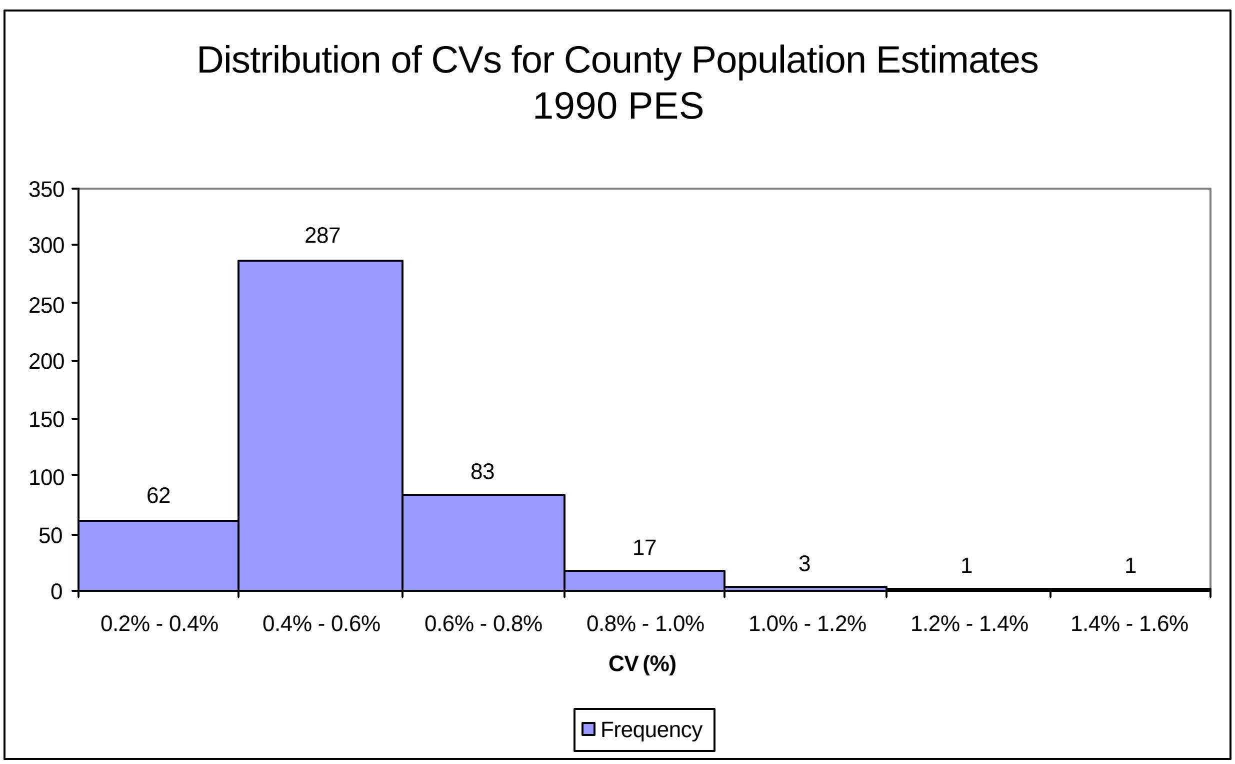
Graph 3b:



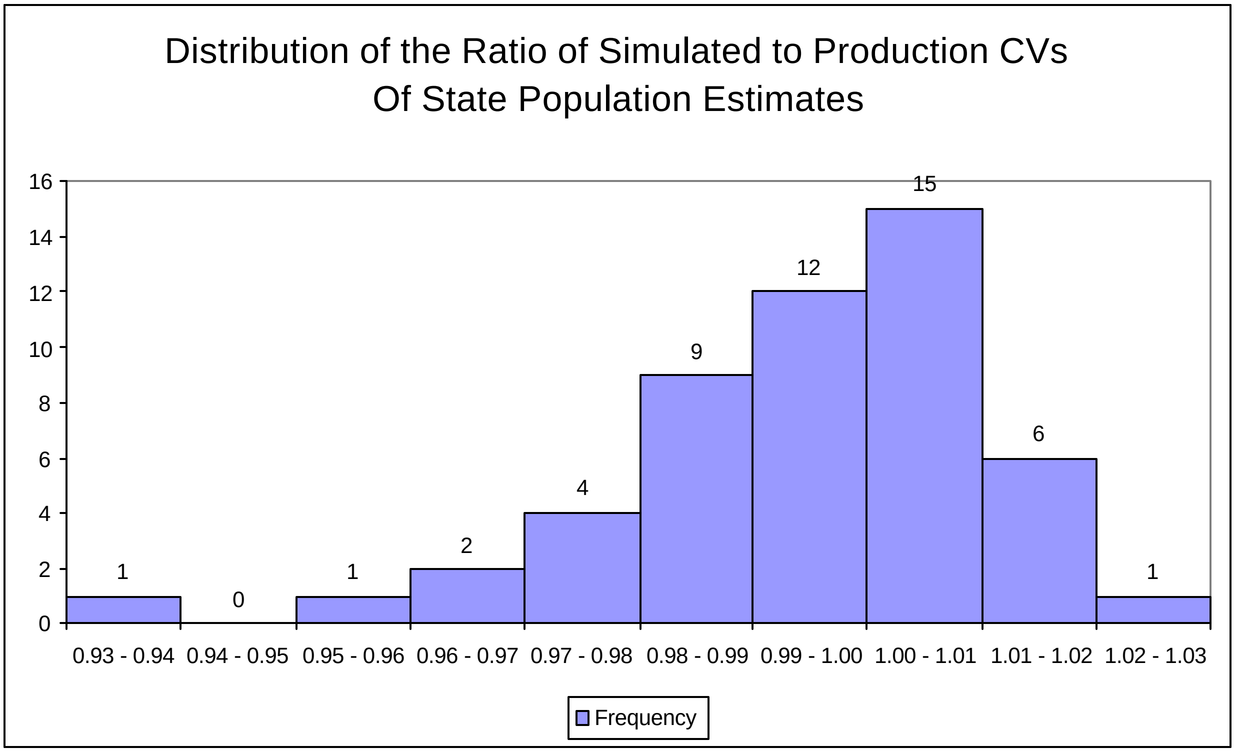
Graph 4a:



Graph 4b:



Graph 5:


Distribution of the Ratio of Simulated to Production CVs
Of State Population Estimates
16
14
12
10
8
6
4
2
0
1
0
1
2
4
9
12
15
6
1
0.93 - 0.94
0.94 - 0.95
0.95 - 0.96
0.96 - 0.97
0.97 - 0.98
0.98 - 0.99
0.99 - 1.00
1.00 - 1.01
1.01 - 1.02
1.02 - 1.03
Frequency

ESCAP MEETING NO. 37 - 02/13/01

MINUTES

Minutes of the Executive Steering Committee on
Accuracy and Coverage Evaluation (A.C.E.) Policy (ESCAP) Meeting # 37

February 13, 2001

Prepared by: Annette Quinlan

The thirty-seventh meeting of the Executive Steering Committee on Accuracy and Coverage Evaluation Policy was held on February 13, 2001 at 10:30. The agenda for the meeting was to discuss A.C.E. Dual System Estimate variances by geographic area and the total error model results.

Committee Attendees:

Nancy Potok
Paula Schneider
Cynthia Clark
Nancy Gordon
John Thompson
Jay Waite
Bob Fay
Howard Hogan
Ruth Ann Killion
John Long
Carol Van Horn

Deputy Director/Acting Director:
William Barron

Other Attendees:

Marvin Raines
Tommy Wright
Donna Kostanich
Raj Singh
William Bell
Deborah Fenstermaker
Mary Mulry
Alfredo Navarro
Michael Starsinic
Nick Birnbaum
Carolee Bush
Kathleen Styles
Maria Urrutia
Sarah Brady
Annette Quinlan

I. A.C.E. Variance Results by Geographic Area

Michael Starsinic presented the coefficients of variation (CV) for various geographic entities, including states, Congressional Districts, and places and counties with populations greater than 100,000 people. Results from 1990 for corresponding geographic entities were also presented for comparison purposes. The graphs distributed at the meeting are attached.

The variance results were reviewed and discussed. The Committee was pleased to note that the graphs show a greater reduction in variance from 1990 to 2000 than was initially anticipated. The reasons for this expected reduction in variance were also discussed and are summarized in the attached document.

The Committee then requested data for additional geographic entities, such as places and counties with populations less than 100,000.

II. Total Error Model Results

It was briefly noted that there are different ways of modeling correlation bias. Mary Mulry then presented results from the Total Error Model Analysis including four treatments of correlation bias. The four different treatments considered are:

- No correlation bias.
- Correlation bias is assumed for Black males but not for Non-black males.
- Correlation bias is assumed for all males except Non-black males between 18-29 years of age.
- Correlation bias is assumed for all groups including18 - 29 year old Non-black males.

It was noted that the intervals of net undercount for minority renters consistently do not include zero, regardless of which treatment is used. The Committee discussed, generally, the significance of the confidence interval touching zero. Those in the Midwest and Northeast, mailout/mailback, low return rate post-stratum grouping continuously showed overcounts for each treatment applied. There was a discussion of identifying plausible causes of these overcounts in the Midwest and Northeast group. John Thompson noted that this finding was consistent with 1990. That is, in 1990, overcounts were also measured in the Northeast. Staff will review more data from the E-sample to further examine any hypothesis regarding the causes, such as the effects of duplicates in the census.

III. Next Meeting

The next meeting scheduled for Wednesday February 14, 2001 will discuss loss functions results and the Census 2000 Full Count Review Program.

ESCAP MEETING NO. 38 - 02/14/01

AGENDA

Kathleen P Porter
02/14/2001 09:04 AM

To: Angela Frazier/DMD/HQ/BOC@BOC, Annette M Quinlan/DMD/HQ/BOC@BOC, Barbara E Hotchkiss/DSD/HQ/BOC@BOC, Betty Ann Saucier/DIR/HQ/BOC@BOC, Carnelle E Sligh/PRED/HQ/BOC@BOC, Carol M Van Horn/DMD/HQ/BOC@BOC, Carolee Bush/DMD/HQ/BOC@BOC, Cynthia Z F Clark/DIR/HQ/BOC@BOC, Deborah A Fenstermaker/DSSD/HQ/BOC@BOC, Donna L Kostanich/DSSD/HQ/BOC@BOC, Hazel V Beaton/SRD/HQ/BOC@BOC, Howard R Hogan/DSSD/HQ/BOC@BOC, John F Long/POP/HQ/BOC@BOC, John H Thompson/DMD/HQ/BOC@BOC, Kathleen M Styles/DMD/HQ/BOC@BOC, Linda A Hiner/DSSD/HQ/BOC@BOC, Lois M Kline/POP/HQ/BOC@BOC, Margaret A Applekamp/DIR/HQ/BOC@BOC, Maria E Urrutia/DMD/HQ/BOC@BOC, Marvin D Raines/DIR/HQ/BOC@BOC, Mary A Cochran/DIR/HQ/BOC@BOC, Mary E Williams/DIR/HQ/BOC@BOC, Nancy A Potok/DIR/HQ/BOC@BOC, Nancy M Gordon/DSD/HQ/BOC@BOC, Nicholas I Birnbaum/DMD/HQ/BOC@BOC, Patricia E Curran/DIR/HQ/BOC@BOC, Paula J Schneider/DIR/HQ/BOC@BOC, Phyllis A Bonnette/DIR/HQ/BOC@BOC, Preston J Waite/DMD/HQ/BOC@BOC, Rajendra P Singh/DSSD/HQ/BOC@BOC, Robert E Fay III/DIR/HQ/BOC@BOC, Ruth Ann Killion/PRED/HQ/BOC@BOC, Sarah E Brady/DMD/HQ/BOC@BOC, Sue A Kent/DMD/HQ/BOC@BOC, Tommy Wright/SRD/HQ/BOC@BOC, Vanessa M Leuthold/DMD/HQ/BOC@BOC, William G Barron Jr/DIR/HQ/BOC@BOC

cc: Alfredo Navarro/DSSD/HQ/BOC@BOC, Michael J Batutis Jr/POP/HQ/BOC@BOC

Subject: Agenda for 2/14 ESCAP

The agenda for the February 14 ESCAP Meeting scheduled from 10:30-12 in Rm. 2412/3 is as follows:

1. Loss Function Results - Freddie Navarro

2. Census Quality: Count Review - Mike Batutis

ESCAP MEETING NO. 38 - 02/14/01

HANDOUTS

Materials attached to these minutes were draft and preliminary material to inform the ESCAP Committee. The data and analysis contained in these documents are subject to revision and are not final. These materials report the results of research and analysis undertaken by Census Bureau staff. They have undergone a more limited review than official Census Bureau publications. Research results and conclusions expressed are those of the authors and do not necessarily indicate concurrence by the Census Bureau.

Demographic Full Count Review

Presentation to ESCAP

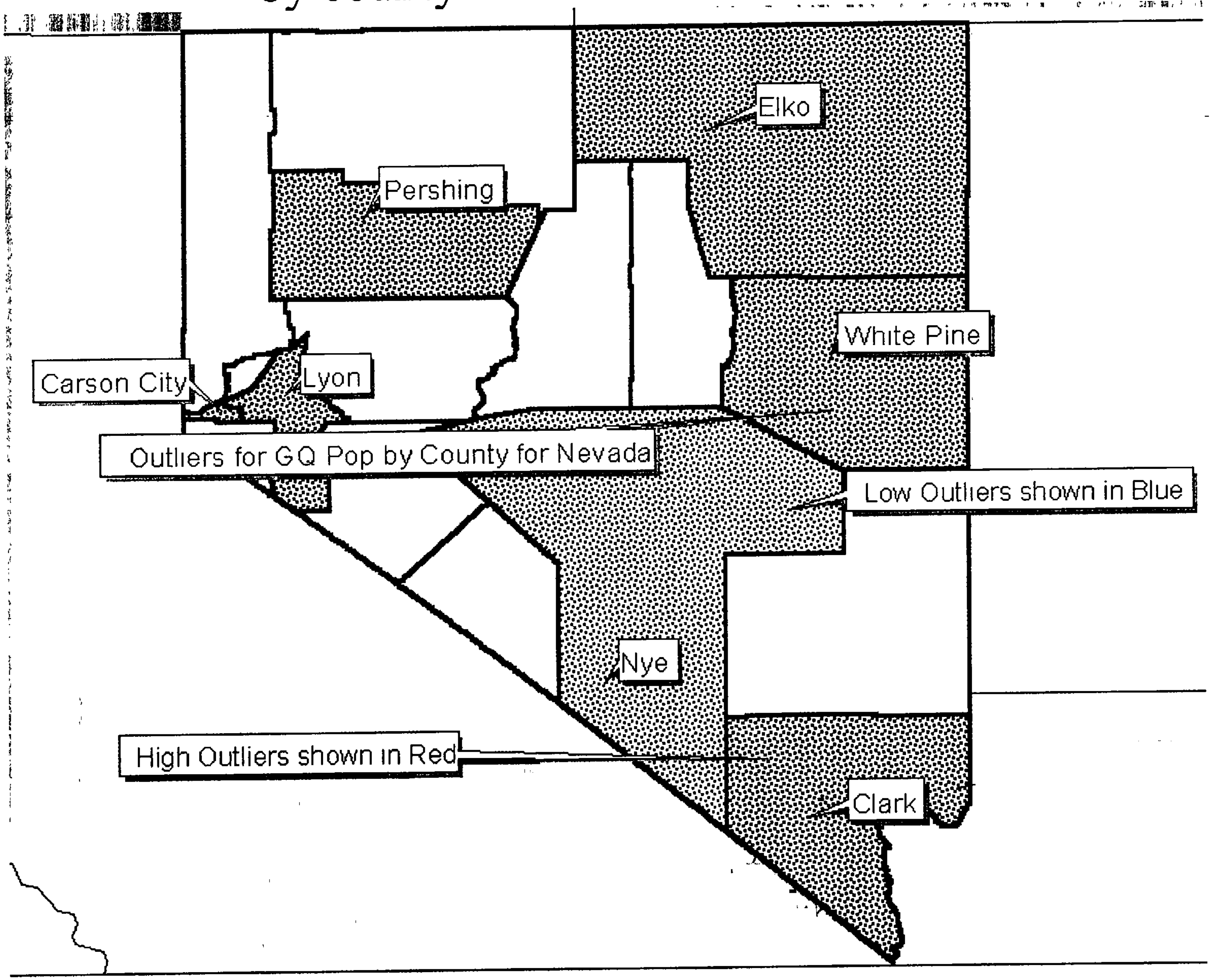
February 14, 2001

Benchmark data

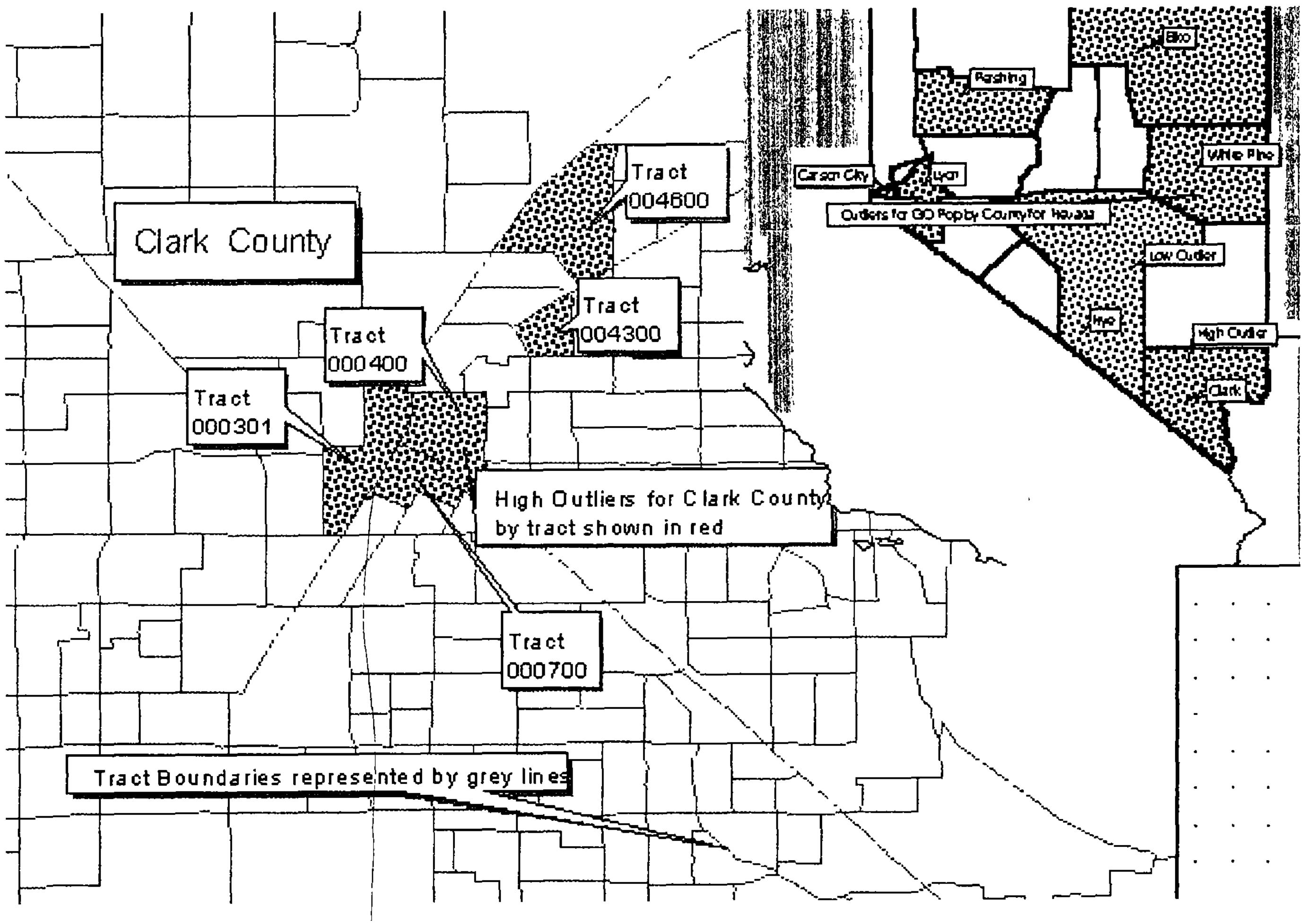
- 1990 Census data at the block level
- Population estimates data extrapolated to April 1, 2000 at the place level*
- Claritas data - Independent, commercially available population estimates for April 1, 2000 at the tract level*

*Not available for Puerto Rico

Analysis example- GQ Pop outliers by county for Nevada



Analysis example- GQ Pop outlier by tract for Clark County, Nevada



GQ Population Counts By Tract For Clark County, Nevada

Comparison of 2000 data with 1990 Census Population Based on HDF Data in **20** Tracts

County Name	County Code	TRACT	[redacted]	GQ2K	GQ90	T_Diff GQ2K_GQ90	P_Diff GQ2K_GQ90	Ratio GQ2K_GQ90	Tolerance pop1990
[redacted]	[redacted]	[redacted]		237	0	237	0%	0	Failed
[redacted]	[redacted]	[redacted]		588	365	223	61.1%	1.61	Failed*
[redacted]	[redacted]	[redacted]		1175	117	1058	904 27%	10 04	Failed*
[redacted]	[redacted]	[redacted]		328	0	328	0%	0	Failed
Clark County	003	[redacted]		1213	0	1213	0%	0	Failed
Clark County	003	000700		2556	1608	948	58 96%	1 59	Failed*
[redacted]	[redacted]	[redacted]	[redacted]	1050	0	1050	0%	0	Failed
[redacted]	[redacted]	[redacted]	[redacted]	200	0	200	0%	0	Failed
[redacted]	[redacted]	[redacted]	[redacted]	808	0	808	0%	0	Failed
[redacted]	[redacted]	[redacted]	[redacted]	203	0	203	0%	0	Failed
[redacted]	[redacted]	[redacted]	[redacted]	207	0	207	0%	0	Failed
[redacted]	[redacted]	[redacted]	[redacted]	657	6	651	10850%	109 5	Failed
[redacted]	[redacted]	[redacted]	[redacted]	573	407	166	40.79%	1 41	Failed
[redacted]	[redacted]	[redacted]	[redacted]	295	28	267	953.57%	10.54	Failed
[redacted]	[redacted]	[redacted]	[redacted]	272	0	272	0%	0	Failed
[redacted]	[redacted]	[redacted]	[redacted]	230	0	230	0%	0	Failed
[redacted]	[redacted]	[redacted]	[redacted]	874	0	874	0%	0	Failed
[redacted]	[redacted]	[redacted]	[redacted]	1765	0	1765	0%	0	Failed
[redacted]	[redacted]	[redacted]		1683	0	1683	0%	0	Failed
[redacted]	[redacted]	[redacted]		540	0	540	0%	0	Failed
Subtotal				**15454**	**2531**				

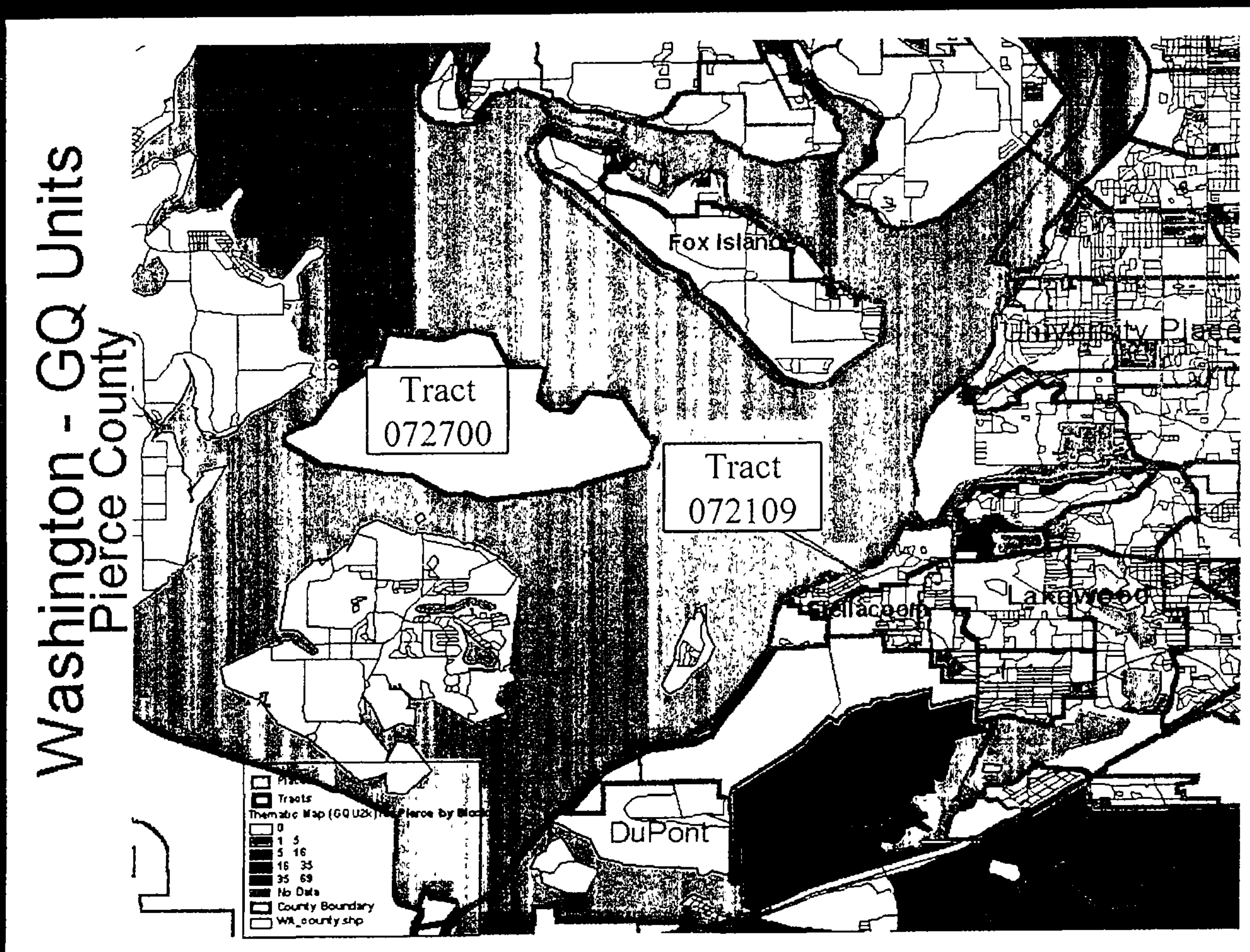
What was accomplished?

- 107 analysts collectively reviewed 252 files and documented 4,330 issues
  - 53 FSCPE reviewers
  - 15 IPC/HHES/POP reviewers
  - 39 Subject matter analysts
- Apportionment count clearance based on thorough review of all state files
- Redistricting data under review for all states

What was learned during Count Review?

- Special Place / Group Quarters (SP/GQ) issues account for 57% of issues documented
- The remaining issues covered general population, Hispanic origin, race, age, sex, and other data
- Documentation of issues does not indicate systematic quality issues
- FSCPE analysts did not identify coverage as a major issue

GQ example: McNeil Island correctional facility, Washington



Tract 072109

Pop2k: 4,427
vs 90: 3,206

GQ Pop2k:
1,469 vs 90: 0

Materials attached to these minutes were draft and preliminary material to inform the ESCAP Committee. The data and analysis contained in these documents are subject to revision and are not final. These materials report the results of research and analysis undertaken by Census Bureau staff. They have undergone a more limited review than official Census Bureau publications. Research results and conclusions expressed are those of the authors and do not necessarily indicate concurrence by the Census Bureau.

Loss Function Analysis Results
Alfredo Navarro, DSSD - 2118-2
February 14, 2001

Purpose of Loss Function Analysis

- To evaluate the accuracy of population counts or shares for the census and the corrected census.

- We compare the census and the corrected census to an estimated "target population" developed from the biases and variance estimates obtained from the Total Error Model. For details see Attachment - Creation of 2000 Error Estimates at the Evaluation PS level.

Unit of Analysis

- Congressional Districts
- States (including the District of Columbia)

What is compared?

- We estimate the difference between the census and the corrected census loss. If the difference is greater than zero, then the A.C.E. is more accurate, otherwise the census is more accurate and adjustment may not be feasible.

Alternative Scenarios of Correlation Bias

- Scenario I - No correlation bias is present in the DSE.
- Scenario II - Correlation bias is present in the DSE for all males except for **Non-black Males 18-29 years of age.**
- Scenario III - Correlation bias is present in the DSE for all males.
- Scenario IV - Correlation bias is present in the DSE for Black males only.

Allocation of Error Components to the A.C.E. Poststrata

- The components of errors estimated at the evaluation poststrata are used to generate the biases for each A.C.E. poststratum.
- Two methods are used to allocate the error to the 416 poststrata:
  GRODSE- The errors are allocated proportional to the size of the DSE.
  GROSUC - The errors are allocated proportional to the size of the net undercount.

Results

- The estimated expected loss for states is lower for the corrected census than for the census for the weighted squared error loss function on population shares for all the targets with and without correlation bias.

- For levels, the expected loss for states is lower for the corrected census than for the census for all loss functions and targets with correlation bias

- The expected loss for congressional districts is lower for the corrected census than for the census for all targets except for the two targets defined by no correlation bias (scenario I). However, in the two cases the percent differences = (D/ACELoss) *100, 0.5 % and 12.3 %, are relatively small.

- For both sets of targets with no correlation bias the difference in the expected loss estimates for states and congressional districts for all loss functions on levels is negative.

Loss Functions

Type of Loss Functions	Census Loss	A.C.E. Loss
1. Squared Error Loss	$\sum_i \left(Cen_i - T_i\right)^2$	$\sum_i \left(ACE_i - T_i\right)^2$
2. Weighted Squared Error Loss	$\sum_i \left(Cen_i - T_i\right)^2 / Cen_i$	$\sum_i \left(ACE_i - T_i\right)^2 / ACE_i$
3. Relative Squared Error Loss	$\sum_i \left(Cen_i - T_i\right)^2 / Cen_i^2$	$\sum_i \left(ACE_i - T_i\right)^2 / ACE_i^2$
4. Equal CD Squared Error Loss (Only for Districts)	$\sum_j Cen_j^2 \sum_i \left(Cen_i - T_i\right)^2$	$\sum_j Cen_j^2 \sum_i \left(ACE_i - T_i\right)^2$

Summary of GRODSE Model

			Square Levels	Weighted Levels	Relative Levels	Square Share	Weighted Share	Relative Share	Equal CD Share
State	Without Correlation Bias	Census Loss	155,696,479,647	8,191.2	0.0013635	0.57566	10.856	609.48	N/A
		ACE Loss	170,491,295,879	15,782.7	0.0035762	0.18509	6.087	628.22	N/A
		Difference	-14,794,816,232	-7,591.5	-0.0022127	0.39057	4.769	-18.73	N/A
	With Correlation Bias except Non-Black 18-29	Census Loss	349,783,046,504	23,107.2	0.0041127	0.48030	13.511	1,142.94	N/A
		ACE Loss	49,051,928,701	5,235.9	0.0017908	0.17774	7.592	966.43	N/A
		Difference	300,731,117,804	17,871.3	0.0023219	0.30256	5.918	176.51	N/A
Congressional District	Without Correlation Bias	Census Loss	7,153,297,039	10,741.6	0.0163355	26.83258	259.619	4,053.11	1,621,053,494
		ACE Loss	12,064,493,231	18,271.6	0.0280708	50.53794	329.841	3,868.69	1,628,821,233
		Difference	-4,911,196,191	-7,530.0	-0.0117353	-23.70537	-70.222	184.42	-7,767,739
	With Correlation Bias except Non-Black 18-29	Census Loss	19,356,711,306	29,884.5	0.0468099	65.72411	717.016	11,011.06	4,093,069,275
		ACE Loss	6,076,602,481	9,152.7	0.0140076	69.99674	508.244	6,201.52	2,478,586,082
		Difference	13,280,108,825	20,731.8	0.0328023	-4.27263	208.773	4,809.54	1,614,483,192

Summary of GROSUC Model

			Square Levels	Weighted Levels	Relative Levels	Square Share	Weighted Share	Relative Share	Equal CD Share
State	Without Correlation Bias	Census Loss	156,084,938,188	7,263.7	0.0007817	0.54927	7.585	218.17	N/A
		ACE Loss	168,765,041,054	16,023.4	0.0040201	0.20906	6.945	800.21	N/A
		Difference	-12,680,102,866	-8,759.7	-0.0032384	0.34021	0.640	-582.04	N/A
	With Correlation Bias except Non-Black 18-29	Census Loss	352,308,245,039	22,382.2	0.0035095	0.44759	10.806	881.97	N/A
		ACE Loss	48,717,072,482	5,616.3	0.0022055	0.19567	8.998	1,264.49	N/A
		Difference	303,591,172,557	16,766.0	0.0013040	0.25192	1.808	-382.52	N/A
Congressional District	Without Correlation Bias	Census Loss	6,390,541,161	9,613.5	0.0146320	16.11157	203.849	3,685.13	1,515,898,059
		ACE Loss	12,424,287,640	18,645.1	0.0284324	66.83943	366.925	4,046.80	1,729,606,494
		Difference	-6,033,746,479	-9,031.6	-0.0138004	-50.72786	-163.076	-361.67	-213,708,435
	With Correlation Bias except Non-Black 18-29	Census Loss	18,681,257,665	28,935.7	0.0454444	57.37441	663.393	10,562.96	3,988,123,703
		ACE Loss	6,477,731,586	9,634.0	0.0145931	88.61298	546.792	6,288.73	2,574,717,192
		Difference	12,203,526,080	19,301.7	0.0308513	-31.23857	116.601	4,274.23	1,413,406,511

Table A. U.S. Summary of Loss Functions Results

Summary of GRODSE Model

			Weighted Levels	Weighted Share	Equal CD Share
State	Without Correlation Bias	Census Loss	8,191.2	10.856	N/A
		ACE Loss	15,782.7	6.087	N/A
		Difference	-7,591.5	4.769	N/A
		Percent Diff	-48.1%	78.3%	N/A
	With Correlation Bias except Non-Black 18-29	Census Loss	23,107.2	13.511	N/A
		ACE Loss	5,235.9	7.592	N/A
		Difference	17,871.3	5.918	N/A
		Percent Diff	341.3%	78.0%	N/A
Congressional District	Without Correlation Bias	Census Loss	N/A	N/A	1,621,053,494
		ACE Loss	N/A	N/A	1,628,821,233
		Difference	N/A	N/A	-7,767,739
		Percent Diff	N/A	N/A	-0.5%
	With Correlation Bias except Non-Black 18-29	Census Loss	N/A	N/A	4,093,069,275
		ACE Loss	N/A	N/A	2,478,586,082
		Difference	N/A	N/A	1,614,483,192
		Percent Diff	N/A	N/A	65.1%

Summary of GROSUC Model

			Weighted Levels	Weighted Share	Equal CD Share
State	Without Correlation Bias	Census Loss	7,263.7	7.585	N/A
		ACE Loss	16,023.4	6.945	N/A
		Difference	-8,759.7	0.640	N/A
		Percent Diff	-54.7%	9.2%	N/A
	With Correlation Bias except Non-Black 18-29	Census Loss	22,382.2	10.806	N/A
		ACE Loss	5,616.3	8.998	N/A
		Difference	16,766.0	1.808	N/A
		Percent Diff	298.5%	20.1%	N/A
Congressional District	Without Correlation Bias	Census Loss	N/A	N/A	1,515,898,059
		ACE Loss	N/A	N/A	1,729,606,494
		Difference	N/A	N/A	-213,708,435
		Percent Diff	N/A	N/A	-12.4%
	With Correlation Bias except Non-Black 18-29	Census Loss	N/A	N/A	3,988,123,703
		ACE Loss	N/A	N/A	2,574,717,192
		Difference	N/A	N/A	1,413,406,511
		Percent Diff	N/A	N/A	54.9%

Summary of GRODSE Model

			Square Levels	Weighted Levels	Relative Levels	Square Share	Weighted Share	Relative Share	Equal CD Share
State	With Correlation Bias in All Groups	Census Loss	260,448,207,267	17,232.5	0.0031715	0.44851	13.450	1,180.52	N/A
		ACE Loss	85,420,366,141	8,208.1	0.0024570	0.18687	8.003	1,005.44	N/A
		Difference	175,027,841,126	9,024.4	0.0007145	0.26163	5.447	175.08	N/A
	With Correlation Bias only for Black Domain	Census Loss	275,890,993,992	17,932.0	0.0032683	0.49950	14.122	1,214.38	N/A
		ACE Loss	78,411,922,184	7,840.0	0.0024093	0.17740	7.804	1,011.52	N/A
		Difference	197,479,071,807	10,092.1	0.0008590	0.32211	6.318	202.86	N/A
Congressional District	With Correlation Bias in All Groups	Census Loss	15,523,029,126	24,141.5	0.0380785	66.58544	737.443	11,309.13	4,167,629,236
		ACE Loss	8,208,649,663	12,323.0	0.0187883	71.61404	530.615	6,552.79	2,602,391,113
		Difference	7,314,379,462	11,818.5	0.0192902	-5.02860	206.828	4,756.34	1,565,238,123
	With Correlation Bias only for Black Domain	Census Loss	16,311,570,953	25,329.1	0.0398900	72.12310	783.554	12,053.15	4,479,025,715
		ACE Loss	7,997,133,553	12,027.6	0.0183745	73.79454	543.412	6,668.29	2,647,580,296
		Difference	8,314,437,400	13,301.5	0.0215155	-1.67144	240.142	5,384.86	1,831,445,419

Summary of GROSUC Model

			Square Levels	Weighted Levels	Relative Levels	Square Share	Weighted Share	Relative Share	Equal CD Share
State	With Correlation Bias in All Groups	Census Loss	262,139,774,103	16,518.9	0.0026326	0.41406	10.781	927.44	N/A
		ACE Loss	84,084,631,464	8,583.1	0.0029267	0.20284	9.435	1,309.76	N/A
		Difference	178,055,142,638	7,935.8	-0.0002941	0.21122	1.346	-382.33	N/A
	With Correlation Bias only for Black Domain	Census Loss	277,816,615,978	17,218.3	0.0027231	0.46853	11.483	962.63	N/A
		ACE Loss	77,517,049,612	8,232.4	0.0028791	0.19699	9.273	1,318.66	N/A
		Difference	200,299,566,366	8,985.9	-0.0001560	0.27154	2.210	-356.03	N/A
Congressional District	With Correlation Bias in All Groups	Census Loss	14,862,358,894	23,190.6	0.0366797	58.04558	682.379	10,837.79	4,053,535,014
		ACE Loss	8,613,229,220	12,783.4	0.0193099	89.88922	567.390	6,615.73	2,688,662,446
		Difference	6,249,129,674	10,407.2	0.0173698	-31.84364	114.989	4,222.06	1,364,872,567
	With Correlation Bias only for Black Domain	Census Loss	15,657,969,033	24,390.7	0.0385127	63.66258	729.601	11,598.15	4,374,628,738
		ACE Loss	8,420,905,021	12,518.4	0.0189444	92.30017	581.568	6,747.66	2,743,863,555
		Difference	7,237,064,012	11,872.3	0.0195683	-28.63759	148.033	4,850.49	1,630,765,182

Table B. U.S. Summary of Loss Functions Results

Summary of GRODSE Model

			Weighted Levels	Weighted Share	Equal CD Share
State	With	Census Loss	17,232.5	13.450	N/A
	Correlation	ACE Loss	8,208.1	8.003	N/A
	Bias in All	Difference	9,024.4	5.447	N/A
	Groups	Percent Diff	109.9%	68.1%	N/A
	With	Census Loss	17,932.0	14.122	N/A
	Correlation	ACE Loss	7,840.0	7.804	N/A
	Bias only for	Difference	10,092.1	6.318	N/A
	Black Domain	Percent Diff	128.7%	81.0%	N/A
Congressional	With	Census Loss	N/A	N/A	4,167,629,236
District	Correlation	ACE Loss	N/A	N/A	2,602,391,113
	Bias in All	Difference	N/A	N/A	1,565,238,123
	Groups	Percent Diff	N/A	N/A	60.1%
	With	Census Loss	N/A	N/A	4,479,025,715
	Correlation	ACE Loss	N/A	N/A	2,647,580,296
	Bias only for	Difference	N/A	N/A	1,831,445,419
	Black Domain	Percent Diff	N/A	N/A	69.2%

Summary of GROSUC Model

			Weighted Levels	Weighted Share	Equal CD Share
State	With	Census Loss	16,518.9	10.781	N/A
	Correlation	ACE Loss	8,583.1	9.435	N/A
	Bias in All	Difference	7,935.8	1.346	N/A
	Groups	Percent Diff	92.5%	14.3%	N/A
	With	Census Loss	17,218.3	11.483	N/A
	Correlation	ACE Loss	8,232.4	9.273	N/A
	Bias only for	Difference	8,985.9	2.210	N/A
	Black Domain	Percent Diff	109.2%	23.8%	N/A
Congressional	With	Census Loss	N/A	N/A	4,053,535,014
District	Correlation	ACE Loss	N/A	N/A	2,688,662,446
	Bias in All	Difference	N/A	N/A	1,364,872,567
	Groups	Percent Diff	N/A	N/A	50.8%
	With	Census Loss	N/A	N/A	4,374,628,738
	Correlation	ACE Loss	N/A	N/A	2,743,863,555
	Bias only for	Difference	N/A	N/A	1,630,765,182
	Black Domain	Percent Diff	N/A	N/A	59.4%

Materials attached to these minutes were draft and preliminary material to inform the ESCAP Committee. The data and analysis contained in these documents are subject to revision and are not final. These materials report the results of research and analysis undertaken by Census Bureau staff. They have undergone a more limited review than official Census Bureau publications. Research results and conclusions expressed are those of the authors and do not necessarily indicate concurrence by the Census Bureau.

Attachment

Creation of 2000 Error Estimates at the Evaluation PS level.

This process involves two major parts:
- Repoststratifying the 1990 data
- Adjusting the results of the 1990 data to reflect design changes, for example the in-scope population definition and the treatment of movers in the P-sample.

To carry out the repostratification, the 1992 MVF file was combined with additional geographic information in order to give each record a near 2000 post-stratum. There are 420 possible poststrata that can be defined given the 1990 census information rather than the 448 possible poststrata which will be defined in Census 2000. The new evaluation poststrata are then collapsings of the 420 (and also of the 448) poststrata down to 16 evaluation poststrata. Defining the evaluation poststrata in this manner allows us to be able to synthetically distribute from the evaluation poststrata to the full 416 poststrata.

The VPLX programs then estimate the error components and covariance matrices using this repoststratified 1992 data. The results are then modified to reflect changes in the definition of the in-scope population , etc. as specified in Mary Mulry's document "Definition of Component Errors for 2000 A.C.E." - Draft 11/6/2000. At this point we now have the 2000 component errors at the evaluation poststrata level. Two differences from 1990 were implemented:
- Several of the denominators were changed where imputations were made from the unresolved cases
- The factor (ddefper/wteper) was included in the covariance matrix in 2000 whereas in 1992 this factor was kept out of the covariance matrix.

Distributing the error components to form bias estimates.

In distributing the bias from the 16 evaluation poststrata to the 416 poststrata, a number of steps must be followed:

1. The gross errors and their associated covariance matrix are first distributed to a set of 112 intermediate poststrata (16 evaluation poststrata by 7 age-sex). The error totals and covariance matrix will also include correlation bias information as appropriate. The correlation bias adds an additional column to the matrix of error totals and adds an additional column and row to the covariance matrix.

2. The output from (1) is then used to construct 1000 simulated DSE's at the intermediate poststratum level which are corrected by the component errors and reflect the covariance structure of the errors.

3. The 1000 simulated DSE's by intermediate poststrata are then distributed to the 416 using two different methods and simulated variable adjustments are made for the ratio estimator bias and for the imputation variance. Two files of 1000 replicates targets by the 416 poststrata are output.

4. Each of the output files are then read and DSE, undercount, and undercount rate is defined for each replicate. Each replicate is also assigned back to an evaluation poststrata.

5. Lastly, a program reads in the 1000 replicates by the 11 evaluation poststrata for the three variables and uses this information to construct an estimate of the target DSE and its variance. Included in the output is a table showing the production undercount, the target undercount, the bias, the sampling variance in the production undercount, the variance in the bias, the total variance, and a 95% confidence interval for the true undercount.

The errors at the evaluation poststrata are distributed to the intermediate poststrata by using the VPLX programs to estimate the gross error components for age-sex crossed with minority / non-minority status (only White or some other race is considered non-minority since American Indians Off-reservation are separated from White domain in 2000). The proportion of error for each age-sex category within minority/non-minority status is calculated and this used to synthetically distribute the error from the evaluation poststrata to the intermediate poststrata. These same proportions are used to create the covariance matrix for the gross errors for each of the intermediate poststrata. It is at the intermediate poststrata level that the correlation bias is incorporated from Bill Bell.

Using the covariance of the gross errors and the correlation bias, 1000 simulated DSEs are created at the intermediate poststratum level. These are then read into the next program which distributes the DSEs from the intermediate poststratum level to the 416 based on two methods:

- Proportion of DSE for each of the 416 poststrata within their respective intermediate poststrata.
- Proportion of Undercount for each of the 416 poststrata within their respective intermediate poststrata. (raw undercount. DSE – Totcenct, not percentage undercount)

Once this is done, then a simulated effect from both the imputations, and ratio-bias is added to the simulated DSEs to obtain the simulated targets. The simulated targets are then divided by the census counts in order to construct a set of simulated adjustment factors. These two files are then saved for the loss function analysis and also for the summary of the total error model analysis.

To produce the summary for the total error analysis, the simulated adjustment factors are used to calculate 1000 simulated DSEs, undercounts and undercount rates for the 416 poststrata. The simulations are then coded for the evaluation poststrata and collapsed to produce three evaluation-level files:

- 1000 simulated DSEs
- 1000 simulated Undercounts
- 1000 simulated Undercount rates

The simulations are then used to calculate the average DSE, Undercount, and Undercount Rate along with their respective variances. The difference between the production values and the averages of the simulations is used to define the bias and the variance of the simulations is used to define the variance of the bias. The total variance is then the sum of the sampling variance of the production values and the variance of the bias. This total variance is then used to construct a 95% confidence interval of the corrected DSE, Undercounts, and Percent Undercounts. This is done twice, once for each distribution method used in creating the 416 poststrata level simulations.

ESCAP MEETING NO. 38 - 02/14/01

MINUTES

Minutes of the Executive Steering Committee on
Accuracy and Coverage Evaluation (A.C.E.) Policy (ESCAP) Meeting # 38

February 14, 2001

Prepared by: Maria Urrutia and Sarah Brady

The thirty-eighth meeting of the Executive Steering Committee on Accuracy and Coverage Evaluation Policy was held on Wednesday, February 14, 2001 at 10:30. The agenda for the meeting was to present Loss Functions results and an overview and findings of the Census 2000 Full Count Review program.

Committee Attendees:

Nancy Potok
Paula Schneider
Cynthia Clark
Nancy Gordon
John Thompson
Jay Waite
Bob Fay
Howard Hogan
Ruth Ann Killion
John Long
Carol Van Horn

Deputy Director/Acting Director:
William Barron

Other Attendees:

Marvin Raines
Tommy Wright
Donna Kostanich
Raj Singh
William Bell
Alfredo Navarro
Deborah Fenstermaker
Michael Batutis
Mary Mulry
Carolee Bush
Nick Birnbaum
Kathleen Styles
Maria Urrutia
Annette Quinlan
Sarah Brady
Roxie Jones
Gretchen Stiers

I. Loss Function Results

Alfredo Navarro presented the results from the loss function analysis, which incorporated the results from the total error model with all error components included. The components of error were based on the estimates of error from 1990 except for correlation bias, sampling error, and ratio estimate bias, which were obtained from the 2000 data. At a previous ESCAP meeting (February 9th, 2001) loss functions were presented that accounted for only sampling error. The purpose of the loss functions from February 9 was to determine if the change between the census and A.C.E. was larger than the sampling error. The loss functions presented at the February 14th meeting estimate the potential improvement of the A.C.E. results as compared to the census.

The loss functions considered several different scenarios for correlation bias. The scenarios were:

- No correlation bias.
- Correlation bias is assumed for Black males but not for Non-black males.
- Correlation bias is assumed for all males except Non-black males between 18-29 years of age.
- Correlation bias is assumed for all groups including 18 - 29 year old Non-black males.

By examining the results of the loss functions for the different correlation bias scenarios, the Committee concluded that correlation bias has a significant effect on both the numeric and distributive loss functions. The results for the equal congressional district share loss functions, where some or all correlation bias was incorporated, indicated that the A.C.E. results were more accurate than the census. In contrast, the equal congressional district share loss function without correlation bias indicated that the census was more accurate than the A.C.E. A discussion was held about the importance of examining alternative formulations to estimate correlation bias, given the influence correlation bias has on loss functions.

The Committee also discussed how to assess the sensitivity of the loss functions to the error parameters that are estimated from 1990 error results, such as matching error. This discussion needs further research and will be continued at a later meeting.

II. Demographic Full Count Review Program

Michael Batutis presented an overview and the findings of the full count review program. The program is summarized in the attached document. The review included such items as total population, Group Quarter (GQ) population, number of GQ units by unit type, and total housing unit population for states, counties, and other small geographic areas.

The majority of the issues documented were for special places/group quarters, which was the initial expectation. The full count review program did not find any serious clustering of errors in the census. Overall, the full count review staff were relatively pleased with the census data reviewed.

III. Next Meeting

The agenda for the next meeting, scheduled for Thursday, February 15, 2001, is to discuss results for the Targeted Extend Search (TES).

ESCAP MEETING NO. 39 - 02/15/01

AGENDA

Kathleen P Porter
02/15/2001 09:25 AM

To: Angela Frazier/DMD/HQ/BOC@BOC, Annette M Quinlan/DMD/HQ/BOC@BOC, Barbara E Hotchkiss/DSD/HQ/BOC@BOC, Betty Ann Saucier/DIR/HQ/BOC@BOC, Carnelle E Sligh/PRED/HQ/BOC@BOC, Carol M Van Horn/DMD/HQ/BOC@BOC, Carolee Bush/DMD/HQ/BOC@BOC, Cynthia Z F Clark/DIR/HQ/BOC@BOC, Deborah A Fenstermaker/DSSD/HQ/BOC@BOC, Donna L Kostanich/DSSD/HQ/BOC@BOC, Hazel V Beaton/SRD/HQ/BOC@BOC, Howard R Hogan/DSSD/HQ/BOC@BOC, John F Long/POP/HQ/BOC@BOC, John H Thompson/DMD/HQ/BOC@BOC, Kathleen M Styles/DMD/HQ/BOC@BOC, Linda A Hiner/DSSD/HQ/BOC@BOC, Lois M Kline/POP/HQ/BOC@BOC, Margaret A Applekamp/DIR/HQ/BOC@BOC, Maria E Urrutia/DMD/HQ/BOC@BOC, Marvin D Raines/DIR/HQ/BOC@BOC, Mary A Cochran/DIR/HQ/BOC@BOC, Mary E Williams/DIR/HQ/BOC@BOC, Nancy A Potok/DIR/HQ/BOC@BOC, Nancy M Gordon/DSD/HQ/BOC@BOC, Nicholas I Birnbaum/DMD/HQ/BOC@BOC, Patricia E Curran/DIR/HQ/BOC@BOC, Paula J Schneider/DIR/HQ/BOC@BOC, Phyllis A Bonnette/DIR/HQ/BOC@BOC, Preston J Waite/DMD/HQ/BOC@BOC, Rajendra P Singh/DSSD/HQ/BOC@BOC, Robert E Fay III/DIR/HQ/BOC@BOC, Ruth Ann Killion/PRED/HQ/BOC@BOC, Sarah E Brady/DMD/HQ/BOC@BOC, Sue A Kent/DMD/HQ/BOC@BOC, Tommy Wright/SRD/HQ/BOC@BOC, Vanessa M Leuthold/DMD/HQ/BOC@BOC, William G Barron Jr/DIR/HQ/BOC@BOC

cc: Douglas B Olson/DSSD/HQ/BOC@BOC

Subject: Agenda for 2/15 ESCAP

The agenda for the February 15 ESCAP Meeting scheduled from 10:30-12 in Rm. 2412/3 is as follows:

1. TES Results - Doug Olson

2. Late Census Adds - Howard Hogan

ESCAP MEETING NO. 39 - 02/15/01

HANDOUTS

Materials attached to these minutes were draft and preliminary material to inform the ESCAP Committee. The data and analysis contained in these documents are subject to revision and are not final. These materials report the results of research and analysis undertaken by Census Bureau staff. They have undergone a more limited review than official Census Bureau publications. Research results and conclusions expressed are those of the authors and do not necessarily indicate concurrence by the Census Bureau.

Effect of the Targeted Extended Search (TES) Operation
Doug Olson and Michael Beaghen, DSSD
February 15, 2001

What is TES?

TES involves field, processing, and matching operations designed to reduce the variance of the Dual-system estimate (DSE) without affecting its expectation. TES accomplished this objective by improving the coverage of the P and E sample by including people in blocks surrounding the A.C.E. clusters.

Why perform TES?

There are geocoding errors of exclusion and inclusion in the sample cluster. Geocoding errors of exclusion affect the P-sample match rate and geocoding errors of inclusion affects the E-sample correct enumeration rate. If the housing unit address is geocoded incorrectly outside the sample cluster, the P-sample people and housing units will not be matched. Conversely, if the housing unit address is geocoded incorrectly in the sample cluster, the E-sample people will be erroneously enumerated. TES reduces erroneous enumeration and non-match rates due to census geocoding error.

In expectation, errors of inclusion and exclusion ought to balance, so the overall measurement of population size would be the same whether TES is performed or not. If the errors balance, TES would have no effect on the estimate of net undercount.

Did TES performed as expected?

Yes. The attached information summarizes the effect of TES on the estimates of correct enumerations and matches at the national level, by regional office, and for subgroups of the population.

Is TES new?

An operation similar to TES, called "Surrounding Block Search", was performed as part of the 1990 PES. The 1990 PES operation differs from TES in that it was performed on all E-sample and P- sample non-matches in all block clusters. The TES operation is "targeted", which means that it directs resources to clusters with high pay-off and to housing units identified as likely to benefit from the extended search.

Why the changes?

Evidence from the 1990 PES evaluation studies suggest that the operation was error prone because the design was inefficient. Anecdotal evidence also suggests that some clerks did not follow the procedures correctly because of the low success rate.

What is the TES design?

The initial Housing Unit Follow-up (February 2000) identified units eligible for TES, these are units that apparently were mis-geocoded. Block clusters that included many of these units were included in TES with certainty. A probability sample was selected from the remaining block clusters with at least potential TES housing unit, that is, a possible geocoding error. About 80 percent of all potential TES housing units were included in the TES sample.

Table I: A.C.E. 2000 -- Effect of TES at the National Level

	With TES (1)	Without TES (2)	Difference (1) - (2)	Effect of TES (1) / (2)
E-sample				
Persons (Ne)	264,578,862	264,634,794	(55,932)	1.000
Corrent Enumerations (CE)	252,096,238	244,387,951	7,708,288	1.032
CE Rate (%)	95.3	92.3	2.93	1.032
P-sample				
Persons (Np)	263,037,259	262,906,916	130,343	1.000
Matches	240,878,622	230,681,205	10,197,418	1.044
Match Rate (%)	91.6	87.7	3.83	1.044
Ratio of CE to Match Rate	1.040	1.053	-0.012	0.989
Standard Error of Ratio (%)	0.134	0.331	-0.197	40.5

Considered at the national level, TES had the expected and intended effects:

- The size of the P- and E-samples are very little changed because the sampling was balanced
- The Match and Correct Enumeration rates increased by roughly similar amounts
- Standard Error was substantially reduced

Table II - Effect of TES by Regional Office

	E-sample CE Rate (%)		P-sample Match %		Ratio CE%/Match%		
	With TES	No TES	With TES	No TES	With TES	No TES	Change (%)
Boston	95.8	93.4	92.0	89.6	1.040	1.042	-0.2
New York	93.3	90.0	88.7	84.4	1.051	1.065	-1.4
Phildelphia	95.5	91.4	91.9	87.2	1.039	1.048	-0.9
Detroit	96.2	93.8	94.0	90.2	1.023	1.040	-1.7
Chicago	95.8	92.8	92.5	89.0	1.036	1.043	-0.7
Kansas City	96.2	94.6	94.1	91.8	1.022	1.031	-0.9
Seattle	95.0	92.6	91.4	87.7	1.039	1.055	-1.6
Charlotte	95.5	91.6	91.3	87.6	1.045	1.045	0
Atlanta	94.6	91.1	90.4	84.1	1.046	1.084	-3.8
Dallas	94.5	91.6	89.9	86.7	1.051	1.057	-0.6
Denver	95.0	92.9	91.4	88.6	1.039	1.048	-0.9
Los Angeles	95.8	92.0	91.1	86.6	1.052	1.062	-1.0

Although there is some expected variation in the effects of TES in different regional offices, it does not appear that any regional office performed TES in a way that distorted its results. TES appears to have been performed similarly in all regions.

Table III -- Effect of TES on Race Domains

	E-sample CE Rate (%)		P-sample Match (%)		Ratio CE%/Match%		
	With TES	No TES	With TES	No TES	With TES	No TES	Change (%)
Indians on Reserv.	95.81	93.24	85.99	77.80	1.114	1.198	-8.4
Indians other	93.97	91.85	87.54	84.77	1.073	1.083	-1.0
Hispanic	94.46	91.84	87.47	83.33	1.080	1.102	-2.2
Black	92.73	89.60	86.94	82.69	1.067	1.084	-1.7
Pacific Islander	93.05	91.01	84.66	81.67	1.099	1.114	-1.5
Asian	94.57	90.47	90.45	86.24	1.046	1.049	-0.3
White/Other	95.90	92.99	93.12	89.43	1.030	1.040	-1.0

The effect of TES is reasonably consistent across demographic groups. The greatest changes are in the smallest population groups.

Table IV -- Effect of TES on Age/Sex Domains

	E-sample CE Rate (%)		P-sample Match (%)		Ratio CE%/Match%		
	With TES	No TES	With TES	No TES	With TES	No TES	Change (%)
AGE / SEX							
0-17	95.94	93.09	90.84	86.97	1.056	1.070	-1.4
M 18-29	92.87	89.61	86.40	82.42	1.075	1.087	-1.2
F 18-29	93.61	89.92	88.54	84.36	1.057	1.066	-0.9
M 30-29	95.23	92.28	91.23	87.53	1.044	1.054	-1.0
F 30-49	96.01	93.04	92.90	89.06	1.033	1.045	-1.2
M 50+	95.34	92.67	93.68	90.03	1.018	1.029	-1.1
F 50+	95.51	92.84	94.29	90.55	1.013	1.025	-1.2

The effect of TES is highly consistent among the age/sex domains.

Table V -- Effect of TES on Race Domains by Tenure

	E-sample CE Rate (%)		P-sample Match %		Ratio CE%/Match%		
	With TES	No TES	With TES	No TES	With TES	No TES	Change (%)
DOMAIN/TENURE							
Indians on Reserv Own	95.65	93.13	85.43	76.62	1.120	1.215	-9.5
Indians on Reserv Rent	96.15	93.48	87.08	80.10	1.104	1.167	-6.3
Indians off Reserv Own	94.55	91.92	89.76	87.03	1.053	1.056	-0.3
Indians off Reserv Rent	93.16	91.76	84.34	81.56	1.105	1.125	-2.0
Hispanic Own	96.25	94.10	90.77	87.38	1.060	1.077	-1.7
Hispanic Rent	92.79	89.75	84.48	79.66	1.098	1.127	-2.9
Black Own	94.25	92.16	90.15	87.06	1.046	1.059	-1.3
Black Rent	91.16	86.96	83.69	78.24	1.089	1.112	-2.3
Pacific Own	93.79	93.05	87.82	85.58	1.068	1.087	-1.9
Pacific Rent	92.33	89.05	82.03	77.70	1.126	1.146	-2.0
Asian Own	95.84	92.96	92.35	88.67	1.038	1.048	-1.0
Asian Rent	92.45	86.39	87.31	82.26	1.059	1.050	0.9
White Own	96.70	94.02	94.59	91.25	1.022	1.030	-0.8
White Rent	93.20	89.56	88.33	83.46	1.055	1.073	-1.8

TES had similar effects on all groups except Indians on Reservations (both owners and renters), a small population group that could be expected to have greater variation. There was a small increase, as opposed to a small decrease, in the CE/Match ratio for Asian renters, another fairly small population group.

Table VI -- Clusters with most E-sample TES Persons

Most E-sample Weighted TES Persons

Cluster	Unweighted Persons	Unweighted Correct Enums	Weighted Persons	Weighted Correct Enums	CE %
213025	107	100	120,463	113,138	93.9
444141	156	150	101,965	97,739	95.9
412437	32	32	100,747	100,747	100.0
622555	78	69	100,136	89,160	89.0
559187	71	69	99,120	95,652	96.5
547935	128	89	88,549	61,569	69.5
341248	22	22	69,175	69,175	100.0
442178	69	64	68,562	63,924	93.2
940486	22	21	66,431	62,974	94.8
541912	78	74	65,336	61,918	94.8
252643	118	114	62,426	60,310	96.6
380725	52	49	58,242	54,412	93.4
372656	20	20	56,023	55,212	98.6
652560	49	49	53,600	53,600	100.0
231498	111	103	53,451	49,807	93.2

Most E-sample Unweighted TES Persons

Cluster	Unweighted Persons	Unweighted Correct Enum	CE %	Weighted Persons	Weighted Correct Enums
550764	321	302	94.0	48,144	45,260
340224	203	198	97.8	9,530	9,317
511220	195	186	95.2	11,929	11,353
331447	179	0	0.0	11,932	0
214403	173	165	95.5	9,200	8,784
923078	172	167	97.2	36,415	35,399
937292	170	114	67.2	36,475	24,505
983502	168	102	60.5	25,925	15,689
380873	162	158	97.6	6,828	6,662
362871	159	143	90.0	43,330	39,000
444141	156	150	95.9	101,965	97,739
374181	152	149	98.1	39,817	39,069
634089	148	144	97.0	43,867	42,569
211623	147	137	92.9	7,818	7,260
441972	143	136	95.3	31,485	29,997

Tables VI and VII (next page) show that in clusters that had the most TES cases the Correct Enumeration and Match rates were very high (in most cases as high as the rates for non-TES persons). TES was very important the clusters where geographic errors were concentrated.

Table VII -- Clusters with most P-sample TES Persons

Most P-sample Weighted TES Persons

Cluster	Unweighted Persons	Unweighted Matches	Weighted Persons	Weighted Matches	Match %
332866	576	526	372,037	343,726	92.4
210237	630	574	181,245	168,798	93.1
926451	778	701	164,381	149,117	90.7
813402	40	38	126,494	120,169	95.0
334292	211	194	124,810	115,897	92.9
940114	245	211	118,511	102,629	86.6
611269	61	46	117,703	88,760	75.4
334722	39	34	111,008	96,849	87.2
993527	38	36	105,286	99,744	94.7
640359	251	205	104,849	86,386	82.4
510974	38	31	104,735	85,442	81.6
546390	329	258	95,030	75,226	79.2
220442	243	205	80,717	70,413	87.2
340711	211	169	71,169	57,011	80.1
391953	21	19	67,352	60,938	90.5

Most P-sample Unweighted TES Persons

Cluster	Unweighted Persons	Unweighted Matches	Unweighted Match %	Weighted Persons	Weighted Matches
926451	778	701	90.1	164,381	149,117
210237	630	574	91.1	181,245	168,798
332866	576	526	91.3	372,037	343,726
546390	329	258	78.4	95,030	75,226
923581	277	223	80.5	57,661	46,637
934612	273	235	86.1	56,552	48,692
362921	252	187	74.2	61,946	46,923
640359	251	205	81.7	104,849	86,386
940114	245	211	86.1	118,511	102,629
220442	243	205	84.4	80,717	70,413
971408	224	0	0.0	17,303	0
334292	211	194	91.9	124,810	115,897
340711	211	169	80.1	71,169	57,011
224865	192	157	81.8	66,210	54,397
225011	181	135	74.6	66,791	49,869

Tables VI (previous page) and VII show that in clusters that had the most TES cases the Correct Enumeration and Match rates were very high (in most cases as high as the rates for non-TES persons). TES was very important the clusters where geographic errors were concentrated.

Table VIII – P and E Sample Surrounding Block Matches

	Count	Weighted
E-sample CE's in surrounding blocks	20,401	7,708,288
P-sample Matches in surrounding blocks	21,878	10,002,072
Adjusted for P-sample coverage		10,676,849

Table IX – Erroneous Enumerations Housing Units

	Count	Weighted
Total EE Housing Units	5,996	1,724,645
With E-sample persons	3,450	1,039,254
No-E-sample persons	2,546	685,391
Persons in EE units	8,104	2,448,863
Correct Enumerations	6,439	1,924,233

The information in tables VIII and IX tell of one possible weakness in the ACE or TES. TES operations added three million more Matches to the P-sample (after coverage adjustment) than Correct Enumerations to the E-sample. Theoretically, these numbers would be equal if all operations were performed perfectly. One possible explanation for part of the difference is shown in Table IX – almost two million Correct Enumeration persons were located in Erroneous Enumeration housing units. It is likely that some of those should have been TES housing units. A follow-up operation, TES2, to determine of there was confounding between Erroneous Enumeration and TES-eligible housing units, is presently underway, but results won't be available for several weeks.

Another possible explanation is that the P-sample listing included some housing units in surrounding blocks. These would have been non-matches had TES *not* been performed, and so TES could have prevented a bias from being introduced. Anecdotal reports from field observations support the likelihood that errors occurred, but no measure of the extent is possible.

Table X

Effect of TES on National CE/Match Ratio

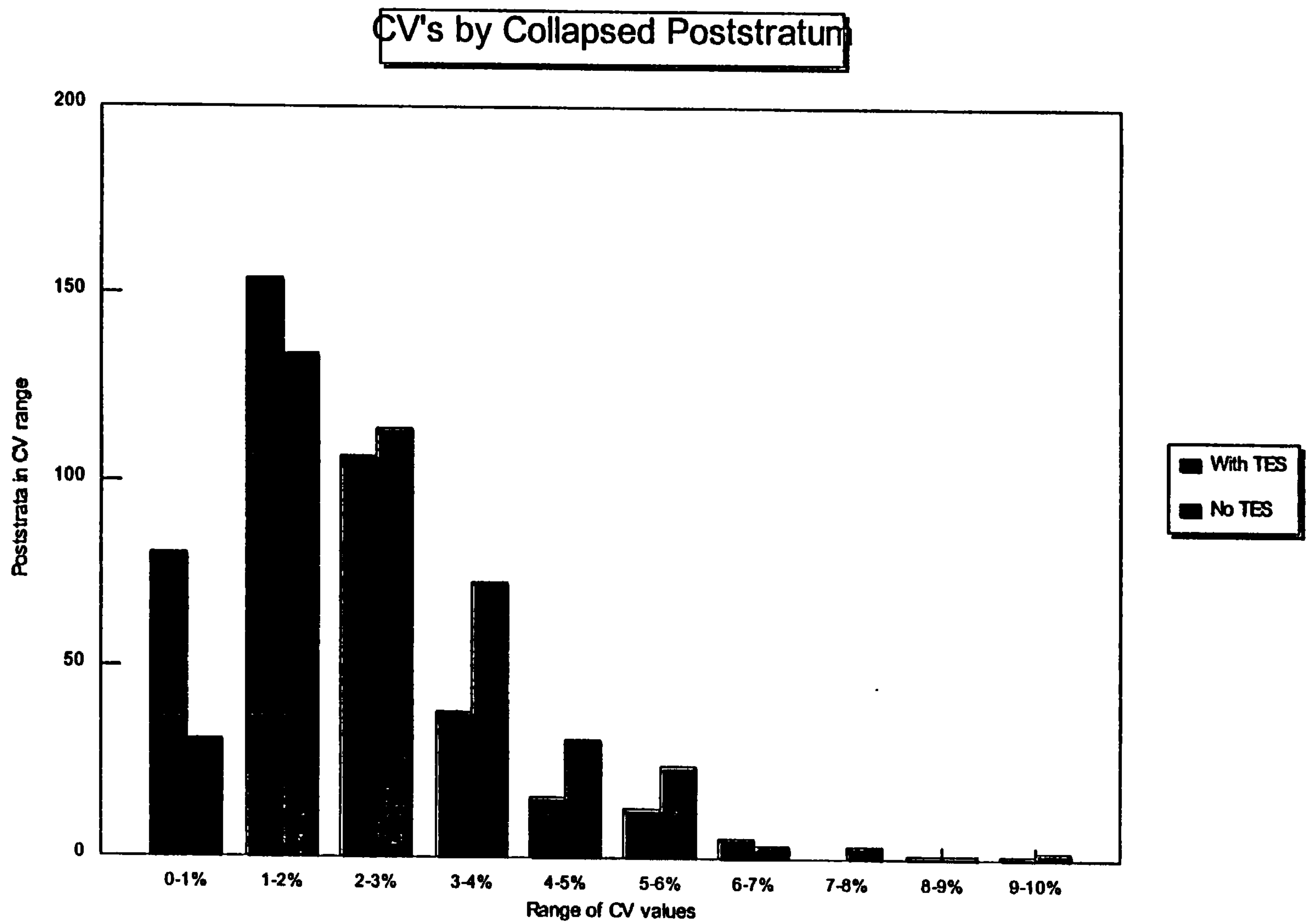
	Ratio CE/Match rate	Std. Err.	CV (%)
TES Performed	1.040	0.13416	0.129
Without TES	1.053	0.33064	0.314

Effect of TES on Poststratum CV's

SE Reduction from TES			Average CV (%)	
At Least	To	Number	TES	No TES
80%	90%	1	1.93	8.79
70%	80%	5	0.91	3.52
60%	70%	18	1.44	3.96
50%	60%	30	1.27	2.73
40%	50%	48	1.43	2.56
30%	40%	56	1.51	2.29
20%	30%	76	1.76	2.30
10%	20%	64	2.22	2.59
0%	10%	42	2.55	2.64
Increase of up to 100%		71	3.12	2.53
Increse more than 100%		5	5.69	2.46
Average CV (%)			2.07	2.66
Median CV (%)			1.81	2.32
Weighted (by Census Count) average CV (%)			1.30	1.93

TES was highly successful in reducing standard errors and consequently coefficients of variation (CV). Although 16% (76 of 416) collapsed poststrata showed increases in standard error, the overall trend is significantly downward. Median and average CV's both increased about 28% under the non-TES assumption. This is similar to the 25% increase estimated for 1990 by Richard Griffiths in a 1995 simulation study, although the two rates are not strictly comparable due to methodological differences.

Graph 1



CV's by Collapsed Poststratum
200
150
100
50
0
Poststrata in CV range
0-1%
1-2%
2-3%
3-4%
4-5%
5-6%
6-7%
7-8%
8-9%
9-10%
Range of CV values
With TES
No TES

Table XI – TES Results by number of TES Housing Units

E-sample

HU/ Clust	Count: Clusters	Count: Units	Count: Persons	Weights: H Units	Weights: Persons	Weights: Cor Enums	CE %
1	178	178	455	228,511	582,421	525,691	90.3
2	71	142	341	162,862	417,673	351,200	84.1
3	50	150	391	173,686	454,612	404,391	89.0
4	47	188	450	128,039	286,271	240,014	83.8
5	33	165	400	134,862	321,583	293,871	91.4
6	25	150	419	105,542	329,542	302,398	91.8
7	25	175	455	84,317	231,633	208,894	90.2
8	21	168	488	80,683	237,220	203,873	85.9
9	31	279	784	189,694	517,630	488,945	94.5
10	15	150	409	72,169	198,341	187,073	94.3
11-20	108	1,608	4,167	653,501	1,654,926	1,563,726	94.5
21-50	132	4,502	9,907	1,171,564	2,516,831	2,217,273	88.1
51	3	153	295	71,430	112,833	101,188	89.7
53	1	53	156	34,642	101,965	97,739	95.9
54	1	54	147	2,872	7,818	7,260	92.9
56	1	56	173	2,978	9,200	8,784	95.5
57	1	57	126	2,049	4,529	929	20.5
58	2	116	260	6,790	14,170	8,668	61.2
59	3	177	377	93,024	182,569	173,767	95.2
60	1	60	195	3,670	11,929	11,353	95.2
62	4	248	502	58,375	116,053	106,684	91.9
63	1	63	75	4,422	5,265	4,919	93.4
64	2	128	154	4,428	5,333	5,258	98.6
66	2	132	169	4,086	5,422	4,907	90.5
67	3	201	398	33,471	59,584	45,442	76.3
68	2	136	318	3,834	10,616	10,357	97.6
69	1	69	94	2,469	3,363	3,345	99.5
70	1	70	148	20,748	43,867	42,569	97.0
73	2	146	279	29,389	53,484	42,503	79.5
95	1	95	321	14,248	48,144	45,260	94.0
Total	768	9,869	22,853	3,578,355	8,544,827	7,708,281	90.2

There does not appear to be any correlation between the number of TES cases in a cluster and the Correct Enumeration rate.

Table XII – TES Results by number of TES Housing Units P-sample

HU/ Cluster	Count: Clusts	Count: Units	Count: Persons	Weights: Housing Units	Weights: Persons	Weights: Matches	Match %
1	351	351	839	516,039	1,217,070	422,108	34.7
2	201	402	968	539,418	1,264,798	581,404	46.0
3	105	315	758	340,998	835,668	375,655	45.0
4	80	320	832	284,378	737,443	389,447	52.8
5	56	280	736	233,023	580,861	367,340	63.2
6	68	408	1,047	241,772	649,090	419,561	64.6
7	48	336	790	220,217	517,232	360,329	69.7
8	40	320	773	160,812	382,209	259,218	67.8
9	39	351	829	179,840	417,749	224,703	53.8
10	26	260	633	169,916	453,345	312,122	68.8
11-20	226	3,289	8,166	1,387,625	3,381,014	2,521,368	74.6
21-50	137	4,377	9,911	1,296,068	2,827,215	1,955,804	69.2
51-100	25	1,598	3,508	548,951	1,034,891	770,214	74.4
103	1	103	251	43,404	104,849	86,386	82.4
106	1	106	277	22,170	57,661	46,637	80.9
134	1	134	211	45,219	71,169	57,011	80.1
139	1	139	576	90,836	372,037	343,726	92.4
156	1	156	211	92,001	124,810	115,897	92.9
214	1	214	329	62,312	95,030	75,226	79.2
319	1	319	630	93,259	181,245	168,798	93.1
386	1	386	778	82,116	164,381	149,117	90.7
Total	1,410	14,164	33,053	6,650,374	15,469,767	10,002,071	64.7

Their appears to be a correlation between the number of TES P-sample housing units and their match rate. Clusters with four or fewer TES units matched at only 44% while those with five or more matched at 72%. At this time, we have no explanation for this difference except pure speculation.

ESCAP MEETING NO. 39 - 02/15/01

MINUTES

Minutes of the Executive Steering Committee on
Accuracy and Coverage Evaluation (A.C.E.) Policy (ESCAP) Meeting # 39

February 15, 2001

Prepared by: Annette Quinlan

The thirty-ninth meeting of the Executive Steering Committee on Accuracy and Coverage Evaluation Policy was held on February 15, 2001 at 10:30. The agenda for the meeting was to discuss results of the A.C.E. Targeted Extended Search operation.

Committee Attendees:

Nancy Potok
Paula Schneider
Cynthia Clark
Nancy Gordon
John Thompson
Jay Waite
Bob Fay
Howard Hogan
Ruth Ann Killion
John Long
Carol Van Horn

Deputy Director/Acting Director:
William Barron

Other Attendees:

Marvin Raines	Danny Childers
Tommy Wright	Doug Olson
Donna Kostanich	Nick Birnbaum
Raj Singh	Sarah Brady
David Whitford	Carolee Bush
Deborah Fenstermaker	Annette Quinlan
Alfredo Navarro	Kathleen Styles
Michael Beaghen	Maria Urrutia

I. Targeted Extended Search (TES)

Howard Hogan began the presentation by illustrating why the issue of balancing the search area between the P and E-samples was so important and describing what we are trying to accomplish through the TES operation.

The TES is designed to aid in reducing the variances associated with the DSEs. In the absence of any A.C.E. geocoding error, the TES would identify about the same number of matches as it would correct enumerations. It was noted that this was not seen in the results presented. In contrast, one would expect the TES to find more P-sample matches than correct enumerations if there was P-sample geocoding error. This may be an explanation of why the results show more P-sample matches than correct enumerations.

Doug Olson characterized the effects of the TES by different post-stratum variables and regional office. The TES results from the Atlanta regional office may appear to be high as compared to other Regional Offices. The Committee has requested the standard errors of these results in order to determine the variation in this number before they decide if its an outlier. There was little evidence of different effects of the TES for most age groups. However, the effect of TES on the race domain of American Indians on Reservations may be the result of P-sample geocoding error.

Danny Childers is conducting a study that consists of investigating housing units where the occupants were classified as correct enumerations, but the housing unit had been classified as erroneous during an earlier operation. In this study, the housing units of the people who were correct enumerations in the census are followed-up to determine if they fall into one of these five categories:

- The housing unit existed in the surrounding blocks.
- The housing unit existed outside the search area.
- The address was not a housing unit.
- The housing unit existed in the cluster.
- The geography was unresolved and no code could be assigned.

It was noted that some of the difference described above can be explained by the A.C.E. having some individuals who were correct enumerations within the A.C.E. cluster but should have been found by the TES in the surrounding blocks. This study also identified categories of individuals who were classified as correct enumerations, either in the A.C.E. cluster or in the surrounding block, but were found to be living more than one block outside the search area, and therefore, should have been coded as an erroneous enumeration. Bob Fay has also expressed some concerns about the TES methodology and is conducting a review. The results of this study will be incorporated into Bob's review.

II. Next Meeting

The next meeting scheduled for Friday February 16, 2001 will discuss results of synthetic estimation, late census adds, and demographic component analysis.

ESCAP MEETING NO. 40 - 02/16/01

AGENDA

Kathleen P Porter
02/15/2001 02:45 PM

To: Angela Frazier/DMD/HQ/BOC@BOC, Annette M Quinlan/DMD/HQ/BOC@BOC, Barbara E Hotchkiss/DSD/HQ/BOC@BOC, Betty Ann Saucier/DIR/HQ/BOC@BOC, Carnelle E Sligh/PRED/HQ/BOC@BOC, Carol M Van Horn/DMD/HQ/BOC@BOC, Carolee Bush/DMD/HQ/BOC@BOC, Cynthia Z F Clark/DIR/HQ/BOC@BOC, Deborah A Fenstermaker/DSSD/HQ/BOC@BOC, Donna L Kostanich/DSSD/HQ/BOC@BOC, Hazel V Beaton/SRD/HQ/BOC@BOC, Howard R Hogan/DSSD/HQ/BOC@BOC, John F Long/POP/HQ/BOC@BOC, John H Thompson/DMD/HQ/BOC@BOC, Kathleen M Styles/DMD/HQ/BOC@BOC, Linda A Hiner/DSSD/HQ/BOC@BOC, Lois M Kline/POP/HQ/BOC@BOC, Margaret A Applekamp/DIR/HQ/BOC@BOC, Maria E Urrutia/DMD/HQ/BOC@BOC, Marvin D Raines/DIR/HQ/BOC@BOC, Mary A Cochran/DIR/HQ/BOC@BOC, Mary E Williams/DIR/HQ/BOC@BOC, Nancy A Potok/DIR/HQ/BOC@BOC, Nancy M Gordon/DSD/HQ/BOC@BOC, Nicholas I Birnbaum/DMD/HQ/BOC@BOC, Patricia E Curran/DIR/HQ/BOC@BOC, Paula J Schneider/DIR/HQ/BOC@BOC, Phyllis A Bonnette/DIR/HQ/BOC@BOC, Preston J Waite/DMD/HQ/BOC@BOC, Rajendra P Singh/DSSD/HQ/BOC@BOC, Robert E Fay III/DIR/HQ/BOC@BOC, Ruth Ann Killion/PRED/HQ/BOC@BOC, Sarah E Brady/DMD/HQ/BOC@BOC, Sue A Kent/DMD/HQ/BOC@BOC, Tommy Wright/SRD/HQ/BOC@BOC, Vanessa M Leuthold/DMD/HQ/BOC@BOC, William G Barron Jr/DIR/HQ/BOC@BOC

cc: Richard A Griffin/DSSD/HQ/BOC@BOC
Subject: Agenda for 2/16 ESCAP

The agenda for the February 16 ESCAP Meeting scheduled from 10:30-12 in Rm. 2412/3 is as follows:

1. Synthetic Error - Rick Griffin

2. Late Census Adds - Howard Hogan

3. Demographic Component Analysis - John Long

ESCAP MEETING NO. 40 - 02/16/01

HANDOUTS

Materials attached to these minutes were draft and preliminary material to inform the ESCAP Committee. The data and analysis contained in these documents are subject to revision and are not final. These materials report the results of research and analysis undertaken by Census Bureau staff. They have undergone a more limited review than official Census Bureau publications. Research results and conclusions expressed are those of the authors and do not necessarily indicate concurrence by the Census Bureau.

February 16, 2001 DRAFT

Executive Summary

We assessed the level of bias in synthetic estimates at the state and congressional district levels. This involved defining the components of synthetic bias, creating artificial populations to estimate one of these components, and estimating the other component by obtaining post-stratum Dual System Estimate levels of bias including correlation bias from the Total Error Model and Loss Function Analysis.

What is the synthetic assumption?

The synthetic assumption holds that census coverage is homogeneous within a particular post-stratum. For example, the synthetic assumption implies that capture probabilities in St. Louis, Missouri in a given post-stratum are the same as capture probabilities in the same post-stratum but in Milwaukee, Wisconsin.

What are synthetic estimates?

A synthetic estimate of population is the sum over post-strata of the post-stratum census coverage correction factor times the post-stratum census count.

What are the components of bias in synthetic estimates?

The bias of a synthetic estimate for a geographic area can be split into two components: (1) synthetic bias due to applying the same coverage correction factor to areas with different census capture probabilities and (2) bias in the Dual System Estimate (DSE) including correlation

How are these components of synthetic bias estimated?

The synthetic bias due to applying the wrong adjustment is estimated using artificial populations. The bias in a synthetic estimate due to DSE is estimated by obtaining the post-stratum level bias in the DSE from the Total Error Model and distributing it to small areas in proportion to their census counts.

What is an artificial population?

We want to compare the synthetic estimates and the census counts for geographic areas with the true counts. However, we do not know the true population for a geographic area such as a congressional district. Surrogate variables correlated with gross undercount and/or gross overcount which are available for small areas are used to create artificial populations. The known population counts for these surrogate variables are scaled to post-stratum level gross undercount and overcount estimates to produce target or true population counts.

OVERVIEW OF METHODOLOGY

This section describes the essence of estimating synthetic bias. There are two components of synthetic bias - synthetic population bias due to applying the same coverage correction factor to areas with different census capture probabilities and bias due to the DSE including correlation bias. The Appendix provides the mathematical details of the methodology.

Creation of Artificial Populations

The basic methodology used to estimate the synthetic population bias component is Artificial Populations.

We use census variables thought to be related to coverage to produce artificial populations. Call these variables surrogates. We use methodology similar to that suggested by Freedman and Wachter (1994, Stat. Sci.). Adjust one surrogate variable to gross undercount and another to gross overcount. These are added and subtracted to census counts to form an artificial population. Unlike the other approaches, this approach can provide both net over- and undercoverage between local areas, within a poststrata. It is possible that the surrogates that are best for gross undercount are different than those that are best for gross overcount.

Surrogate variables considered:

- Allocations: households with more than a specified amount of item nonresponse (items: race, Hispanic Origin. relationship, sex, and age)
- # Non-Mail Returns (May be good proxy for gross overcount)
- #Substitutions whole-household imputes and/or partial household substitutions (May be good proxy for gross undercount)
- # duplicates added back (late adds) (gross overcount?)
- -#duplicates not added back (gross undercount?)

At the A.C.E block cluster level, within post-strata, one can construct an indicator of total coverage, the coverage gap, as follows:

z=(P_sample count - matches)- (E_sample count- correct enumerations).

<u>Block Cluster Level Analysis:</u> At the block cluster level, a correlation between z and each artificial population's true values can be made. Note that each artificial population uses two surrogate variables, one for gross undercount and one for gross overcount. Because of the, possibly, large amount of geocoding error at the block cluster level, these correlations will likely be small, or large correlations may merely mean that our artificial populations are related to geocoding error. However, they may be used to help rank the artificial populations in order of importance.

From this analysis multiple sets of artificial populations are selected for calculation of the bias of synthetic estimates.

Bias of Synthetic Estimator

The bias of a synthetic estimate can be split into two components:

- synthetic population bias due to applying the same coverage correction factor to areas with different census capture probabilities
- bias in the DSE including correlation bias.

The first component is estimated using artificial populations, the second component is estimated using post-stratum biases, estimated as part of the Total Error Model and Loss Function work.

Results

What are the Results of the Artificial Population Creation?

Based on the block cluster level correlation analysis four artificial populations were created as described in Table 1.

Table 1: Surrogate Variables used to Create Artificial Populations

	Undercount Surrogate	Overcount Surrogate
Artificial Population 1	(#non-GQ persons) - (#persons in whole household substitutions)	(#non-GQ persons) - (# persons for whom date of birth was allocated consistent with reported age)
Artificial Population 2	(#non-GQ persons) - (#persons in whole household substitutions)	(#non-GQ persons) - (#persons in whole household substitutions)
Artificial Population 3	# non-GQ persons with 2 or more item allocations	# persons for whom date of birth was allocated consistent with reported age
Artificial Population 4	# non-GQ persons whose household did not mail back the questionnaire	# non-GQ persons whose household did not mail back the questionnaire

Note that for Artificial Populations 2 and 4 the same surrogate variable is used for undercount and overcount.

Regional Examples of Artificial Population Creation

Tables 2 and 3 below illustrate the creation of the first two artificial population counts at the regional level. The actual artificial populations are created at the congressional district level and summed to the state and region levels. Thus these illustrations are not exactly equal to what is obtained by summing over the congressional districts but they are very close.

For each table the total U.S. gross undercount is allocated to the regions in proportion to their totals for the undercount surrogate variable. The total U.S. gross overcount is allocated to the regions in proportion to their totals for the overcount surrogate variable. The artificial population count is then given by: census count + allocated gross undercount - allocated gross overcount.

Table 2 : Illustration of Artificial Population 1 Creation at Regional Level

region	census count (1)	undercount surrogate (2)	overcount surrogate (3)	allocated undercount (4)	allocated overcount (5)	artificial pop. count (1)-(4)+(5)
northeast	51,926,613.00	51,149,463.00	50,402,601.00	3,085,402.28	2,473,932.51	52,538,082.77
Midwest	62,600,946.00	62,010,357.00	61,063,365.00	3,740,545.56	2,997,199.36	63,344,292.19
South	97,400,148.00	96,112,343.00	94,599,821.00	5,797,621.80	4,643,283.63	98,554,486.17
West	61,659,290.00	60,874,702.00	59,476,763.00	3,672,041.37	2,919,323.50	62,412,007.87
Total	273,586,997.00	270,146,865.00	265,542,550.00			
U.S. gross undercount		16,295,611.00				
U.S. gross overcount		13,033,739.00				

Table 3 : Illustration of Artificial Population 2 Creation at Regional Level

region	census count (1)	undercount surrogate (2)	overcount surrogate (3)	allocated undercount (4)	allocated overcount (5)	artificial pop. count (1)-(4)+(5)
northeast	51,926,613.00	51,149,463.00	51,149,463.00	3,085,402.28	2,467,801.17	52,544,214.10
Midwest	62,600,946.00	62,010,357.00	62,010,357.00	3,740,545.56	2,991,805.25	63,349,686.31
South	97,400,148.00	96,112,343.00	96,112,343.00	5,797,621.80	4,637,119.12	98,560,650.68
West	61,659,290.00	60,874,702.00	60,874,702.00	3,672,041.37	2,937,013.46	62,394,317.91
Total	273,586,997.00	270,146,865.00	270,146,865.00			
U.S. gross undercount		16,295,611.00				
U.S. gross overcount		13,033,739.00				

At the state level, using the artificial populations how does the total bias in the synthetic estimates compare with the bias in the census numbers?

For a given state let absolute census bias be defined as the absolute value of the census count (or share) minus the true count (or share) from the artificial population. Similarly let the absolute synthetic bias be defined as the absolute value of the synthetic estimate of count (or share) minus the true count (or share) from the artificial population. Next define the ratio, R, of the absolute census bias to the absolute synthetic bias.

$$R = \frac{|census - true|}{|synthetic - true|}$$

Tables 4 and 5 show the percentiles of the ratio R for the artificial populations. At the tails of the distributions of the ratios for shares, the values are quite small (or large) because the census (or the synthetic estimate) is very close to the true value as measured by the artificial population.

Table 4: Percentiles of ratios of Absolute Census Bias to Absolute Synthetic Bias- Artificial Populations 1 and 2

Percentile	Count	Count	Share	Share
	Artificial Population 1	Artificial Population 2	Artificial Population 1	Artificial Population 2
5	0.525	0.530	0.165	0.182
10	0.745	0.718	0.398	0.427
25	1.12	1.13	0.971	0.889
50	1.50	1.52	1.99	2.38
75	2.06	2.07	6.74	7.68
90	2.61	2.49	10.76	14.50
95	2.89	2.91	28.82	23.97

Table 5: Percentiles of ratios of Absolute Census Bias to Absolute Synthetic Bias- Artificial Populations 3 and 4

Percentile	Count	Count	Share	Share
	Artificial Population 3	Artificial Population 4	Artificial Population 3	Artificial Population 4
5	0.067	0.310	0.078	0.312
10	0.228	0.562	0.190	0.536
25	0.439	1.13	0.540	0.895
50	1.04	1.53	1.20	1.99
75	3.61	2.19	2.10	5.60
90	10.71	3.1	11.11	14.47
95	11.18	4.13	23.54	26.39

At the congressional district level, how does the total bias in the synthetic estimates compare with the bias in the census numbers?

Tables 6 and 7 show the percentiles of the ratio R for the artificial populations. At the tails of the distributions of the ratios for shares, the values are quite small (or large) because the census (or the synthetic estimate) is very close to the true value as measured by the artificial population.

Table 6: Percentiles of ratios of Absolute Census Bias to Absolute Synthetic Bias- Artificial Populations 1 and 2

Percentile	Count	Count	Share	Share
	Artificial Population 1	Artificial Population 2	Artificial Population 1	Artificial Population 2
5	0.426	0.410	0.074	0.115
10	0.662	0.650	0.189	0.232
25	1.06	1.08	0.723	0.766
50	1.56	1.57	2.13	2.06
75	2.32	2.30	4.78	4.66
90	3.53	3.36	12.81	11.48
95	4.16	3.87	26.29	25.50

Table 7: Percentiles of ratios of Absolute Census Bias to Absolute Synthetic Bias- Artificial Populations 3 and 4

Percentile	Count	Count	Share	Share
	Artificial Population 3	**Artificial Population 4**	**Artificial Population 3**	**Artificial Population 4**
5	**0.073**	**0.278**	**0.147**	**0.095**
10	**0.141**	**0.524**	**0.328**	**0.230**
25	**0.399**	**0.928**	**0.665**	**0.597**
50	**1.40**	**1.62**	**1.44**	**1.60**
75	**3.99**	**2.72**	**2.96**	**3.73**
90	**11.53**	**5.66**	**6.40**	**7.68**
95	**22.75**	**11.05**	**12.65**	**18.68**

What are levels of the components of Synthetic Bias for states?

Tables 8, 9, 10, and 11 give the components of synthetic bias at the State level for Artificial Populations 1, 2, 3 and 4 and 2 respectively. Columns (1) through (4) are for estimated of count. Column (1), SPB, is the synthetic population bias and column (2), SCB, is the DSE level bias including correlation bias. Column (3) is the percentage of total bias that comes from SPB. Column (4) is the relative total bias in the state level synthetic estimate of the count. Column (5) is the bias in the estimate of share. Column (6) is the relative bias in the synthetic estimate of population share.

Table 8: State Level Synthetic Bias Using Artificial Population 1

State	Synthetic Bias (1) SPB	DSE Bias (2) SCB	Percent Synthetic Bias (3) SPB / (SPB+SCB)	Rel. Bias of Count (4) (SPB+SCB)/N	Bias of Share (5) B-share	Rel. Bias of Share (6) rel. B-share
Alabama	1394.94	40533.60	3.33%	0.0096	0 000037	0.0023
Alaska	-337.36	5756.61	-6.23%	0.0096	0.000005	0.0023
Arizona	412.98	34663.24	1.18%	0.0069	-0.000006	-0.0003
Arkansas	-481.13	27496.00	-1.78%	0.0103	0.000029	0 0030
California	-10067.30	189121.83	-5 62%	0.0053	-0.000227	-0.0019
Colorado	-464.19	27683.67	-1.71%	0.0064	-0.000012	-0.0008
Connecticut	-727.94	21182.84	-3.56%	0.0061	-0.000013	-0.0011
Delaware	475.17	4482.92	9.58%	0.0064	-0.000002	-0.0008
D.C.	147.90	3615.50	3.93%	0.0069	-0.000001	-0.0003
Flordia	-1482.94	91291.25	-1.65%	0.0057	-0.000087	-0.0015
Georgia	-413.56	67633.99	-0.62%	0.0083	0.000032	0.0011
Hawaii	145.03	8936.81	1.60%	0.0076	0.000001	0.0003
Idaho	-89.67	10219.72	-0.89%	0.0079	0.000003	0.0007
Illinois	2563.32	92868.06	2.69%	0.0078	0.000027	0.0006
Indiana	2568.49	44217.50	5.49%	0.0079	0.000014	0.0006
Iowa	-415.77	24639.43	-1.72%	0.0085	0.000013	0.0013
Kansas	-381.65	22185.29	-1.75%	0.0083	0.000010	0.0011
Kentucky	-377.36	33810.55	-1.13%	0.0084	0.000017	0.0012
Louisiana	124.39	39117.80	0.32%	0.0089	0.000027	0.0017
Maine	-420.31	11978.16	-3.64%	0.0092	0.000009	0.0020
Maryland	2895.27	28527.04	9.21%	0.0060	-0.000023	-0.0012
Massachusetts	-1404.17	39611.87	-3.68%	0.0062	-0.000023	-0.0010
Michigan	-2227.99	61938.80	-3.73%	0.0061	-0.000038	-0.0011
Minnesota	-193.06	34760.24	-0.56%	0.0072	-0.000000	-0.0000
Mississippi	-385.25	30977.12	-1.26%	0.0110	0.000038	0.0037
Missouri	-843.03	43858.30	-1.96%	0.0079	0.000013	0.0007
Montana	-484.12	10844.03	-4.67%	0.0116	0.000014	0.0044
Nebraska	-297.82	13349.36	-2.28%	0.0078	0.000004	0.0006
Nevada	715.30	15252.22	4.48%	0.0080	0.000006	0.0008
New Hampshire	315.20	8933.91	3.41%	0.0076	0.000002	0.0004
New Jersey	-927.77	54644.22	-1.73%	0.0065	-0.000023	-0.0008
New Mexico	920.05	19416.28	4.52%	0.0112	0.000026	0.0039
New York	10838.10	158133.60	6.41%	0.0091	0.000125	0.0019
North Carolina	-1227.27	71076.73	-1.76%	0.0088	0.000046	0.0016
North Dakota	-6.01	6391.76	-0.09%	0.0103	0.000007	0.0030
Ohio	-3561.09	67350.23	-5.58%	0.0057	-0.000059	-0.0015
Oklahoma	-1139.75	30577.96	-3.87%	0.0087	0.000018	0.0015
Oregon	-698.72	17415.06	-4.18%	0.0049	-0.000028	-0.0023
Pennsylvania	-72.79	83376.19	-0.09%	0.0070	-0.000010	-0.0002
Rhode Island	428.66	7348.05	5.51%	0.0076	0.000002	0.0004
South Carolina	294.51	29295.11	1.00%	0.0075	0.000005	0.0003
South Dakota	-0.61	7800.69	-0.01%	0.0107	0.000009	0.0034
Tennessee	191.04	40271.95	0.47%	0.0072	-0.000000	-0.0000
Texas	6887.96	161274.77	4.10%	0.0081	0.000069	0.0009
Utah	-1095.79	13784.96	-8.64%	0.0057	-0.000012	-0.0015
Vermont	14.60	5697.71	0.26%	0.0096	0.000005	0.0023
Virginia	518 06	47846.52	1.07%	0.0070	-0.000006	-0.0002
Washington	-1711.25	26599.31	-6.88%	0.0043	-0.000062	-0.0029
West Virginia	-739.03	20328.15	-3.77%	0.0110	0.000024	0.0037
Wisconsin	675.89	34693.46	1.91%	0.0067	-0.000009	-0.0005
Wyoming	147.82	4988.61	2.88%	0 0105	0.000006	0.0033
Average				0.0079		0.0007
Standard Deviation				0.0017		0.0017

Table 9: State Level Synthetic Bias Using Artificial Population 2

State	Synthetic Bias (1) SPB	DSE Bias (2) SCB	Percent Synthetic Bias (3) SPB / (SPB+SCB)	Rel. Bias of Count (4) (SPB+SCB)/N	Bias of Share (5) B-share	Rel. Bias of Share (6) rel. B-share
Alabama	199.54	40533.60	0.49%	0.0093	0.000033	0.0021
Alaska	-141.68	5756 61	-2.52%	0.0099	0 000005	0.0027
Arizona	455.69	34663.24	1.30%	0.0069	-0.000006	-0.0003
Arkansas	-189.99	27496.00	-0.70%	0.0104	0.000030	0.0031
California	-1971.65	189121.83	-1.05%	0.0056	-0.000197	-0.0016
Colorado	177.16	27683.67	0.64%	0.0066	-0.000010	-0.0007
Connecticut	-76.51	21182.84	-0.36%	0.0063	-0.000010	-0.0009
Delaware	147 41	4482.92	3.18%	0.0060	-0.000003	-0.0012
D.C.	46.23	3615.50	1.26%	0.0067	-0.000001	-0.0005
Flordia	318.93	91291.25	0.35%	0.0058	-0.000080	-0.0014
Georgia	119.81	67633.99	0.18%	0.0084	0.000034	0.0012
Hawaii	-19.78	8936.81	-0.22%	0.0074	0.000001	0.0002
Idaho	-35.84	10219.72	-0.35%	0.0079	0.000003	0.0007
Illinois	723.59	92868.06	0.77%	0.0077	0.000020	0.0005
Indiana	457.74	44217.50	1.02%	0.0075	0.000006	0.0003
Iowa	-81.59	24639.43	-0.33%	0.0087	0.000015	0.0014
Kansas	-94.80	22185.29	-0.43%	0.0084	0.000011	0.0012
Kentucky	-260.21	33810.55	-0.78%	0.0084	0.000017	0.0012
Louisiana	-114.65	39117.80	-0.29%	0.0089	0.000026	0.0017
Maine	-61.45	11978.16	-0.52%	0.0095	0.000010	0.0022
Maryland	646.81	28527.04	2.22%	0.0056	-0.000031	-0.0016
Massachusetts	-74.56	39611.87	-0.19%	0.0064	-0.000018	-0.0008
Michigan	-404.92	61938.80	-0.66%	0.0063	-0.000032	-0.0009
Minnesota	-70.25	34760.24	-0.20%	0.0072	0.000000	0.0000
Mississippi	-189.90	30977.12	-0.62%	0.0111	0.000038	0.0038
Missouri	-219.52	43858.30	-0.50%	0.0080	0.000015	0.0008
Montana	-88.55	10844.03	-0.82%	0.0121	0.000016	0.0048
Nebraska	-90.61	13349.36	-0.68%	0.0079	0.000004	0.0007
Nevada	194.42	15252.22	1.26%	0.0077	0.000004	0.0005
New Hampshire	102.63	8933.91	1.14%	0.0074	0.000001	0.0002
New Jersey	-7.49	54644.22	-0.01%	0.0066	-0.000019	-0.0006
New Mexico	139.92	19416.28	0.72%	0.0108	0.000023	0.0035
New York	2144.24	158133.60	1.34%	0.0086	0.000094	0.0014
North Carolina	-391.92	71076.73	-0.55%	0.0089	0.000049	0.0017
North Dakota	-31.17	6391.76	-0.49%	0.0102	0.000007	0.0030
Ohio	-681.49	67350.23	-1.02%	0.0060	-0.000049	-0.0012
Oklahoma	-354.30	30577.96	-1.17%	0.0089	0.000021	0.0017
Oregon	-92.01	17415.06	-0.53%	0.0051	-0.000026	-0.0021
Pennsylvania	-29.00	83376.19	-0.03%	0.0070	-0.000010	-0.0002
Rhode Island	145.62	7348.05	1.94%	0.0074	0.000001	0.0001
South Carolina	37.37	29295.11	0.13%	0.0075	0.000004	0.0003
South Dakota	-29.14	7800.69	-0.37%	0.0106	0.000009	0.0034
Tennessee	-111.02	40271.95	-0.28%	0.0072	-0.000001	-0.0001
Texas	522.60	161274.77	0.32%	0.0078	0.000046	0.0006
Utah	-258.48	13784.96	-1.91%	0.0061	-0.000009	-0.0011
Vermont	42.31	5697.71	0.74%	0.0096	0.000005	0.0024
Virginia	-122.22	47846.52	-0.26%	0.0069	-0.000008	-0.0003
Washington	-195.31	26599.31	-0.74%	0.0045	-0.000056	-0.0027
West Virginia	-160.66	20328.15	-0.80%	0.0113	0.000026	0.0040
Wisconsin	2.96	34693.46	0.01%	0.0066	-0.000011	-0.0006
Wyoming	25.72	4988 61	0.51%	0.0103	0.000005	0.0031
Average				0.0079		0.0007
Standard Deviation				0.0017		0.0017

Table 10: State Level Synthetic Bias Using Artificial Population 3

State	Synthetic Bias (1) SPB	DSE Bias (2) SCB	Percent Synthetic Bias (3) SPB / (SPB+SCB)	Rel Bias of Count (4) (SPB+SCB)/N	Bias of Share (5) B-share	Rel. Bias of Share (6) rel. B-share
Alabama	-20428.94	40533 60	-101.61%	0.0046	-0.000042	-0.0026
Alaska	9432.73	5756.61	62 10%	0.0273	0.000040	0.0199
Arizona	-26995.03	34663.24	-352 04%	0 0015	-0.000105	-0.0057
Arkansas	7222.78	27496.00	20.80%	0 0132	0.000057	0.0060
California	-65950.08	189121.83	-53.54%	0.0037	-0.000428	-0.0035
Colorado	-1859.27	27683.67	-7.20%	0 0061	-0.000017	-0.0011
Connecticut	3556.02	21182.84	14.37%	0.0074	0.000003	0.0002
Delaware	-6138.60	4482.92	370.76%	-0.0021	-0.000026	-0.0093
D.C.	-10760.92	3615.50	150.60%	-0.0128	-0.000040	-0.0199
Flordia	-46967.64	91291.25	-105.97%	0.0028	-0.000251	-0.0044
Georgia	-78860.97	67633.99	702.42%	-0.0014	-0.000251	-0.0085
Hawaii	7676.49	8936.81	46.21%	0.0139	0.000029	0.0066
Idaho	6100.35	10219.72	37.38%	0.0128	0.000025	0.0055
Illinois	-47306.73	92868.06	-103.83%	0.0037	-0.000154	-0.0035
Indiana	-24874.84	44217.50	-128.60%	0.0032	-0.000085	-0.0039
Iowa	6621.27	24639.43	21.18%	0.0110	0.000039	0.0038
Kansas	-713.31	22185.29	-3.32%	0.0082	0.000009	0.0010
Kentucky	29131.67	33810.55	46.28%	0.0159	0.000124	0.0087
Louisiana	18482.77	39117.80	32.09%	0.0132	0.000093	0.0059
Maine	-9615.13	11978.16	-406.90%	0.0019	-0.000024	-0.0053
Maryland	-6434.16	28527.04	-29.12%	0.0042	-0.000056	-0.0030
Massachusetts	2785.78	39611.87	6.57%	0.0069	-0.000008	-0.0003
Michigan	3880.59	61938.80	5.90%	0.0067	-0.000016	-0.0005
Minnesota	-162.86	34760.24	-0.47%	0.0072	-0.000000	-0.0000
Mississippi	-10475.56	30977.12	-51.10%	0.0073	0.000001	0.0001
Missouri	18607.08	43858.30	29.79%	0.0115	0.000083	0.0042
Montana	6478.82	10844.03	37.40%	0.0196	0.000039	0.0123
Nebraska	11078.14	13349.36	45.35%	0.0147	0.000045	0.0075
Nevada	-6370.91	15252.22	-71.73%	0.0044	-0.000020	-0.0028
New Hampshire	-6872.91	8933.91	-333.47%	0.0017	-0.000024	-0.0055
New Jersey	2588.22	54644.22	4.52%	0.0069	-0.000010	-0.0003
New Mexico	7435.74	19416.28	27.69%	0.0148	0.000049	0.0076
New York	-1846.14	158133.60	-1.18%	0.0084	0.000079	0.0012
North Carolina	21199.96	71076.73	22.97%	0.0117	0.000127	0.0045
North Dakota	3358.77	6391.76	34.45%	0.0158	0.000019	0.0085
Ohio	48030.82	67350.23	41.63%	0.0104	0.000127	0.0032
Oklahoma	24028.31	30577.96	44.00%	0.0162	0.000109	0.0090
Oregon	17523.54	17415.06	50.16%	0.0104	0.000038	0.0031
Pennsylvania	-23313.35	83376.19	-38.81%	0.0050	-0.000094	-0.0022
Rhode Island	-1067.70	7348.05	-17.00%	0.0062	-0.000004	-0.0010
South Carolina	15319.84	29295.11	34.34%	0.0114	0.000059	0.0042
South Dakota	3724.95	7800.69	32.32%	0.0159	0.000023	0.0086
Tennessee	17912.91	40271.95	30.79%	0.0104	0.000064	0.0032
Texas	37130.92	161274.77	18.71%	0.0096	0.000178	0.0024
Utah	13426.96	13784.96	49.34%	0.0123	0.000040	0.0051
Vermont	-4460.12	5697.71	-360.39%	0.0021	-0.000011	-0.0051
Virginia	61915.38	47846.52	56.41%	0.0160	0.000216	0.0087
Washington	-17646.87	26599.31	-197.12%	0.0015	-0.000119	-0.0056
West Virginia	8534.49	20328.15	29.57%	0.0162	0.000057	0.0090
Wisconsin	4386.53	34693.46	11.22%	0.0075	0.000005	0.0002
Wyoming	1550.19	4988.61	23.71%	0.0135	0.000011	0.0062
Average				0.0086		0.0014
Standard Deviation				0 0065		0.0065

Table 11: State Level Synthetic Bias Using Artificial Population 4

State	Synthetic Bias (1) SPB	DSE Bias (2) SCB	Percent Synthetic Bias (3) SPB / (SPB+SCB)	Rel. Bias of Count (4) (SPB+SCB)/N	Bias of Share (5) B-share	Rel Bias of Share (6) rel. B-share
Alabama	-3236.31	40533.60	-8 68%	0.0085	0.000020	0 0013
Alaska	2770.83	5756 61	32.49%	0.0151	0.000016	0 0079
Arizona	-5669.03	34663.24	-19.55%	0.0057	-0.000028	-0.0015
Arkansas	-1199.67	27496.00	-4.56%	0.0100	0.000026	0.0027
California	20832.05	189121.83	9.92%	0.0063	-0.000115	-0.0009
Colorado	-355.08	27683.67	-1.30%	0.0064	-0.000012	-0.0008
Connecticut	1339.26	21182.84	5.95%	0.0068	-0.000005	-0.0004
Delaware	-1873.52	4482.92	-71.80%	0.0034	-0.000011	-0.0038
D.C.	2999.11	3615.50	45.34%	0.0121	0.000010	0.0049
Flordia	-21247.13	91291.25	-30.33%	0.0044	-0.000158	-0.0028
Georgia	7514.80	67633.99	10.00%	0.0093	0.000061	0.0021
Hawaii	2334.08	8936.81	20.71%	0.0094	0.000009	0.0022
Idaho	-408.65	10219.72	-4.17%	0.0076	0.000002	0.0004
Illinois	-46239.74	92868.06	-99.17%	0.0038	-0.000150	-0.0034
Indiana	-16104.48	44217.50	-57.28%	0.0047	-0.000054	-0.0025
Iowa	1142.13	24639.43	4.43%	0.0091	0.000019	0.0019
Kansas	-3612.48	22185.29	-19.45%	0.0071	-0.000001	-0.0001
Kentucky	360.38	33810.55	1.05%	0.0086	0.000020	0.0014
Louisiana	-6757.10	39117.80	-20.88%	0.0074	0.000002	0.0001
Maine	-2478.42	11978.16	-26.09%	0.0075	0.000001	0.0003
Maryland	-3017.18	28527.04	-11.83%	0.0049	-0.000044	-0.0023
Massachusetts	7210.68	39611.87	15.40%	0.0076	0.000008	0.0004
Michigan	-3273.46	61938.80	-5.58%	0.0060	-0.000042	-0.0012
Minnesota	2303.40	34760.24	6.21%	0.0077	0.000009	0.0005
Mississippi	1668.00	30977.12	5.11%	0.0117	0.000045	0.0045
Missouri	-4967.88	43858.30	-12.77%	0.0071	-0.000002	-0.0001
Montana	1393.98	10844.03	11.39%	0.0138	0.000021	0.0065
Nebraska	-75.65	13349.36	-0.57%	0.0079	0.000004	0.0007
Nevada	2002.80	15252.22	11.61%	0.0086	0.000010	0.0014
New Hampshire	-2702.62	8933.91	-43.37%	0.0051	-0.000009	-0.0021
New Jersey	10763.37	54644.22	16.46%	0.0079	0.000020	0.0007
New Mexico	-226.57	19416.28	-1.18%	0.0106	0.000022	0.0033
New York	72414.66	158133.60	31.41%	0.0124	0.000347	0.0052
North Carolina	-10190.45	71076.73	-16.74%	0.0077	0.000014	0.0005
North Dakota	3073.96	6391.76	32.47%	0.0153	0.000018	0.0080
Ohio	-10849.74	67350.23	-19.20%	0.0051	-0.000085	-0.0021
Oklahoma	-1463.17	30577.96	-5.03%	0.0086	0.000017	0.0014
Oregon	-8686.88	17415.06	-99.53%	0.0026	-0.000057	-0.0046
Pennsylvania	8681.66	83376.19	9.43%	0.0077	0.000021	0.0005
Rhode Island	-393.97	7348.05	-5.67%	0.0068	-0.000001	-0.0004
South Carolina	-5429.62	29295.11	-22.75%	0.0061	-0.000016	-0.0011
South Dakota	1867.10	7800.69	19.31%	0.0133	0.000016	0.0060
Tennessee	-9880.13	40271.95	-32.51%	0.0054	-0.000036	-0.0018
Texas	17829.74	161274.77	9.95%	0.0087	0.000108	0.0015
Utah	268.50	13784.96	1.91%	0.0063	-0.000007	-0.0009
Vermont	-2539.49	5697.71	-80.41%	0.0053	-0.000004	-0.0019
Virginia	7661.62	47846.52	13.80%	0.0080	0.000020	0.0008
Washington	-5345.54	26599.31	-25.15%	0.0036	-0.000075	-0.0036
West Virginia	7295.36	20328.15	26.41%	0.0155	0.000053	0.0083
Wisconsin	-4038.41	34693.46	-13.17%	0.0058	-0.000026	-0.0014
Wyoming	-1465.12	4988.61	-41.58%	0 0072	-0.000000	-0.0000
Average				0.0079		0.0007
Standard Deviation				0.0030		0.0030

APPENDIX

I. Forming artificial populations

Let X denote a surrogate for gross undercount and Y denote a surrogate for gross overcount.

DSE_j = the Dual System Estimate for Post-stratum j

E_j = the weighted E sample total in post-stratum j

CE_j = the weighted E sample number of correct enumerations in post-stratum j

EE_j = the weighted E sample number of erroneous enumerations in post-stratum j

$Cen_{.j}$ = the census count in post-stratum j

Note that for any variable V, $V_{.j}$ is the sum of V_{ij} over areas i.

Define the estimated gross undercount as follows:

$$GUNDER_j = DSE_j - Cen_{.j}\left(\frac{CE_j}{E_j}\right)$$

Define the estimated gross overcount as follows:

$$GOVER_j = Cen_{.j}\left(\frac{EE_j}{E_j}\right)$$

N_{ij} is the artificial population value and Cen_{ij} is the census count for area i, post-stratum j.

$$N_{ij} = Cen_{ij} + X_{ij}\frac{GUNDER_j}{X_{.j}} - Y_{ij}\frac{GOVER_j}{Y_j}$$

$$N_{.j} = Cen_{.j} + GUNDER_j - GOVER_j = Cen_{.j} + DSE_j - Cen_{.j} = DSE_j$$

II. The estimate of bias for area i takes the following form:

$$B_i = SPB_i + SCB_i = (\hat{N}_{i.} - N_{i.}) + \sum_j \frac{Cen_{ij}}{Cen_{.j}}\hat{D}_j.$$

Here, the first part is estimated from an artificial population; it is the artificial population synthetic count (equivalent to the production synthetic estimate because the artificial populations are adjusted so that the total over areas for a post-stratum equals the DSE) minus the actual population count from the artificial population.

The second part contains the post-stratum bias, $\hat{D}_j$, (estimated elsewhere) which is an estimate of: (E(DSE$_j$)-true the population of post-stratum j). In this second term, we weight the post-stratum bias by the proportion of post-stratum census counts in area i .

III. The bias for the synthetic estimator of a population share for area i takes the following form:

$$B_{share,\, i} = \frac{N_i + SPB_i + SCB_i}{\sum_i (N_i + SPB_i + SCB_i)} - \frac{N_i}{\sum_i N_i}$$

January 17, 2001 DRAFT

Synthetic Error

How do we form synthetic estimates?

- Form estimates of the true population count using Dual System Estimation (DSE) within each post-stratum.
- Calculate Coverage Correction Factors (CCF).
- Apply CCFs to post-stratum counts down to the block level.

What is synthetic error?

- Synthetic error is introduced when net undercount is **not** homogeneous for geographic areas within a post-stratum.
- The model does not require exact homogeneity within each post-stratum; only that all areas within a post-stratum have similar undercount without too much variation.
- Lack of homogeneity effects census counts as well as adjusted counts.
- By definition, synthetic error does not exist at the aggregate post-stratum level and does not contribute to errors in the DSE. (e.g. if a small town is part of only one post-stratum, the town could have synthetic error).

How does synthetic error compare to errors in the DSE?

- Recall that DSE errors include sampling error, correlation bias, matching error, data error and error in adjusting for missing information.
- At the post-stratum level and higher all the error is due to the DSE since synthetic error is zero at these levels.
- Synthetic error is expected to become relatively more important than DSE error as the geographic area becomes smaller.
- For very small areas like blocks, the synthetic error is expected to be the dominate source of error.

What do we know about synthetic error?

- If the census count for an area has a large local error (e.g. many housing units were mis-geocoded or whole apartment buildings were counted twice), then the area can also have a large synthetic error.

- It is difficult to measure, particularly for small areas.

How can synthetic error be measured?

- Develop an estimate based on data from that area alone. In practice there is not enough sample in small areas so these estimates have too much variance.

- Develop a model using a surrogate variable with a known geographic distribution. This is referred to as "Artificial Populations".

What is the strategy of the analysis to deal with synthetic error?

- Demonstrate that synthetic error will not reverse decisions from comparing the loss due to adjustment with the loss due to the census.

- In 1990, the loss function analysis was not seriously distorted in favor of adjustment due to synthetic error.

Our primary concern is synthetic error for Congressional Districts. There is not sufficient A.C.E. sample in each Congressional District to produce an estimate from data in that area alone. Consequently, the "Artificial Populations" methodology will be used to get a rough estimate of synthetic error at the Congressional District level.

Consider the following oversimplification:

- There is only 1 post-stratum.
- There is no sampling error.
- There are no biases in the DSE.
- There is undercount in the census.
- Synthetic error is present.

The dual system estimate results are:

DSE = 22,000
CCF = 1.073

Assume we know the true count for 5 areas comprising the post-stratum.

Table 1. Example - Synthethic Error Only

Area	*True**	Census	Syn. Est	*Cen. Error**	*Syn. Error**
1	*10,000*	9,000	9,658.5	*1000*	*341.5*
2	*5,000*	4,500	4,829.3	*500*	*170.7*
3	*3,500*	3,000	3,219.5	*500*	*280.5*
4	*2,500*	2,000	2,146.3	*500*	*353.7*
5	*1,000*	2,000	2,146.3	*-1000*	*-1146.3*
PS Total	*22,000*	20,500	22,000.0	*1500*	*0.0*

** Truth is obviously not observable.*

1. Since there is no error in the DSE the DSE is equal to the true count.

2. The coverage correction factor (CCF) is the DSE divided by the census count for the post-stratum.

3. Note that there is no synthetic error for all areas combined at the post-stratum level.

4. If the census error is large, the synthetic error can also be large.

How is synthetic error measured using artificial populations?

Using the same scenario given in the above example and the distribution of whole person substitutions by geographic area, consider the following:

Table 2. Example - Measuring Synthethic Error w/Artificial Population

Area	*True**	Census	Syn. Est	*Cen. Error**	*Syn. Error**	Whole Person Substitution	Estimated True	Estimated Syn. Error
1	*10,000*	9,000	9,658.5	*1000*	*341.5*	10	9,150.0	-508.5
2	*5,000*	4,500	4,829.3	*500*	*170.7*	30	4,950.0	120.7
3	*3,500*	3,000	3,219.5	*500*	*280.5*	30	3,450.0	230.5
4	*2,500*	2,000	2,146.3	*500*	*353.7*	30	2,450.0	303.7
5	*1,000*	2,000	2,146.3	*-1000*	*-1146.3*	0	2,000.0	-146.3
PS Total	*22,000*	20,500	22,000.0	*1500*	*-0.0*	100	22,000.0	0.0

** Truth is obviously not observable.*

DSE = 22,000
CCF = 1.073
UC = 1,500

1. The distribution of whole person substitutions is used to estimate the true count. This is done by distributing the undercount proportional to the substitutions.

2. The estimated synthetic error is the difference between the estimated truth and the synthetic estimate.

3. Note that estimating synthetic error can be problematic.

Our actual artificial population construction will adjust one surrogate variable to gross undercount and another to gross overcount. These are added and subtracted to census counts to form an artificial population. This is illustrated in the following example:

Table 3. Example - Measuring Synthethic Error w/Artificial Population
Gross Undercount and Overcount Correction

Area	*True**	Census	Syn. Est	Whole Person Substitution	Persons in Late Adds	Gross Undercount Correction	Gross Overcount Correction	Estimated True	Estimated Syn. Error
1	*10,000*	9,000	9,658.5	10	10	300	-300	9,000.0	-658.5
2	*5,000*	4,500	4,829.3	30	10	900	-300	5,100.0	270.7
3	*3,500*	3,000	3,219.5	30	5	900	-150	3,750.0	530.5
4	*2,500*	2,000	2,146.3	30	0	900	0	2,900.0	753.7
5	*1,000*	2,000	2,146.3	0	25	0	-750	1,250.0	-896.3
PS Total	*22,000*	20,500	22,000.0	100	50	3,000	(1,500)	22,000.0	0.0

** Truth is obviously not observable.*

1. The distributions of whole person substitutions is used to estimate the gross undercount. This is done by distributing the gross undercount proportional to the whole person substitutions.

2. The distributions of persons in late adds is used to estimate the gross overcount. This is done by distributing the gross overcount proportional to persons in late adds.

3. The true count is then estimated by adding the gross undercount and subtracting the gross overcount from the census count.

4. This approach can provide both net overcoverage and undercoverage within a post-stratum.

5. The following surrogate variables will be used to create alternative artificial populations: census whole person substitutions, census allocations of data items, single or multi-unit address, mail return rate, duplicates reinstated (late adds), duplicates not reinstated.

6. We will use A.C.E. block cluster data to choose combinations of surrogate variables for gross undercount and gross overcount that are correlated with block cluster level net coverage error.

How does synthetic bias effect comparison of synthetic estimates with census counts?

- Loss function analysis will compare the loss of using census counts for local areas to the loss of using synthetic estimates using a target estimate derived using a similar synthetic assumption.

- We will use artificial populations to provide information about bias resulting from using a synthetic population target instead of the true population target.

ESCAP MEETING NO. 40 - 02/16/01

MINUTES

Minutes of the Executive Steering Committee on Accuracy and Coverage Evaluation (A.C.E.) Policy (ESCAP) Meeting # 40
February 16, 2001

Prepared by: Nick Birnbaum

The fortieth meeting of the Executive Steering Committee on Accuracy and Coverage Evaluation Policy was held on February 16, 2001 at 10:30.

The agenda for the meeting was to discuss the methodology for the synthetic bias analysis and the results.

Committee Attendees:

Paula Schneider
Cynthia Clark
Nancy Gordon
John Thompson
Jay Waite
Howard Hogan
Ruth Ann Killion
John Long
Carol Van Horn

Other Attendees:

Marvin Raines	Richard Griffin
Tommy Wright	Nick Birnbaum
Donna Kostanich	Sarah Brady
Raj Singh	Carolee Bush
William Bell	Annette Quinlan
Donald Malec	Kathleen Styles
Deborah Fenstermaker	Maria Urrutia

I. Synthetic Estimation and Associated Error

DSSD staff provided background information on synthetic estimation and associated error. Synthetic estimation is the process by which the coverage correction factors from the Dual System Estimates (DSEs) are carried down to the block level. The synthetic assumption states that the people in a particular post-stratum are relatively homogeneous and will generally share the same coverage factor. Synthetic error is introduced when net undercount is not homogeneous for geographic areas within a post-stratum. It is expected to become relatively more important than components of the DSE error as the geographic area becomes smaller. For very small areas like blocks, the synthetic error is expected to be the dominant source of error. However, lack of homogeneity affects census counts for very small areas, like blocks, as well because of considerable variation in net coverage rates; for example, when a large multi-unit structure is geocoded into the wrong block.

The estimate of synthetic error in the adjusted counts below the post-stratum level, is based on data from geographic levels pertinent to our analysis. While the congressional district is the level of relevance to the Committee's decision, there is not sufficient A.C.E. sample in each congressional district to produce an estimate from data in those areas alone. Therefore, artificial populations are created using surrogate variables with known distributions for the areas of analysis. Surrogate variables correlated with gross undercount and gross overcount, which are available for the areas of analysis, are used to create the artificial populations. The known population counts for these surrogate variables are scaled to post-stratum level gross undercount and overcount estimates to produce target or true population counts. Synthetic error for the areas of analysis can then be calculated as the difference between the target populations and the synthetic estimates. This method of analysis is a significant improvement over that used in 1990 to estimate synthetic population bias. In the 1990 analysis, only net undercount was allocated for artificial populations, but in 2000, both net overcount and net undercount are used to allocate the artificial populations.

DSSD staff then presented information and data on the analyses they have conducted. To estimate synthetic population bias at the state and congressional district levels, they examined a number of potential surrogate or indicator variables (including the number of allocations, the number of non-mail returns, the number of substitutions, etc.) at the A.C.E. block cluster level to determine how well they correlated with a rough indicator of net coverage at the A.C.E. block cluster level. They identified four artificial populations each containing a set of surrogate variables for undercount and overcount.

Two analyses were presented to the Committee. In the first analysis the following ratios of census bias to synthetic bias were considered. See page 5 of the attached document for a definition of the ratio. These ratios were considered for counts and shares for states and Congressional Districts. For states the distribution of the ratios indicated that the synthetic estimate of count (and share) was an improvement over the census count (and share) more often than not. For congressional districts, the synthetic estimate of count and share showed improvement over the census count and share for all four artificial populations. The importance of this analysis demonstrates that we do not have a situation where only a small proportion of the areas are improved, while the majority are dis-improved.

The second analysis examined the synthetic bias relative to the DSE bias (or the bias measured by the Total Error Model) for the four artificial populations at the state level. For two of the artificial populations, the synthetic population bias was a relatively small component of the total bias. However, for the other two artificial populations, it was a fairly large component. Consequently, the Committee determined that it would be important to assess the effect of synthetic bias on the loss function analyses. That is, synthetic error effects both the adjusted and unadjusted census accuracy or loss. Therefore, it is important to study the relative increase or decrease to the difference between the census and adjusted census loss.

II. Next Meeting

The agenda for the next meeting, to be held on February 19, 2001, is to examine the effect of reinstated cases on the Dual System Estimates and the Coverage Correction Factors, discuss revised Demographic Analysis estimates, and to identify outstanding issues for the Committee's upcoming deliberations.

ESCAP MEETING NO. 41 - 02/19/01

AGENDA

Kathleen P Porter
02/15/2001 11:32 AM

To: Angela Frazier/DMD/HQ/BOC@BOC, Annette M Quinlan/DMD/HQ/BOC@BOC, Barbara E Hotchkiss/DSD/HQ/BOC@BOC, Betty Ann Saucier/DIR/HQ/BOC@BOC, Carnelle E Sligh/PRED/HQ/BOC@BOC, Carol M Van Horn/DMD/HQ/BOC@BOC, Carolee Bush/DMD/HQ/BOC@BOC, Cynthia Z F Clark/DIR/HQ/BOC@BOC, Deborah A Fenstermaker/DSSD/HQ/BOC@BOC, Donna L Kostanich/DSSD/HQ/BOC@BOC, Howard R Hogan/DSSD/HQ/BOC@BOC, John F Long/POP/HQ/BOC@BOC, John H Thompson/DMD/HQ/BOC@BOC, Kathleen M Styles/DMD/HQ/BOC@BOC, Linda A Hiner/DSSD/HQ/BOC@BOC, Lois M Kline/POP/HQ/BOC@BOC, Margaret A Applekamp/DIR/HQ/BOC@BOC, Maria E Urrutia/DMD/HQ/BOC@BOC, Marvin D Raines/DIR/HQ/BOC@BOC, Mary A Cochran/DIR/HQ/BOC@BOC, Mary E Williams/DIR/HQ/BOC@BOC, Nancy A Potok/DIR/HQ/BOC@BOC, Nancy M Gordon/DSD/HQ/BOC@BOC, Nicholas I Birnbaum/DMD/HQ/BOC@BOC, Patricia E Curran/DIR/HQ/BOC@BOC, Paula J Schneider/DIR/HQ/BOC@BOC, Phyllis A Bonnette/DIR/HQ/BOC@BOC, Preston J Waite/DMD/HQ/BOC@BOC, Rajendra P Singh/DSSD/HQ/BOC@BOC, Robert E Fay III/DIR/HQ/BOC@BOC, Ruth Ann Killion/PRED/HQ/BOC@BOC, Sarah E Brady/DMD/HQ/BOC@BOC, Sue A Kent/DMD/HQ/BOC@BOC, Vanessa M Leuthold/DMD/HQ/BOC@BOC, William G Barron Jr/DIR/HQ/BOC@BOC

cc:

Subject: ESCAP Meeting for Feb. 19

The ESCAP meeting on Monday February 19 will begin at 10:30 in Rm. 2412/3. The attendees will be:

Bill Barron
Nancy Potok
Paula Schneider
John Thompson
Jay Waite
Nancy Gordon
Cynthia Clark
Ruth Ann Killion
Carol M. Van Horn
Howard Hogan
John Long
Bob Fay
Marvin Raines

Technical support staff on call:

Donna Kostanich
Raj Singh
Debbie Fenstermaker

ESCAP MEETING NO. 41 - 02/19/01

HANDOUTS

Materials attached to these minutes were draft and preliminary material to inform the ESCAP Committee. The data and analysis contained in these documents are subject to revision and are not final. These materials report the results of research and analysis undertaken by Census Bureau staff. They have undergone a more limited review than official Census Bureau publications. Research results and conclusions expressed are those of the authors and do not necessarily indicate concurrence by the Census Bureau.

UNITED STATES DEPARTMENT OF COMMEI
Bureau of the Census
Washington, DC 20233-0001

May 6, 1996

MEMORANDUM FOR Ruth Ann Killion
Chief, Decennial Statistical Studies Division

From: C Leggieri for
Working Group on the Use of Demographic Analysis in Census 2000

Subject: Working Group Report

This transmits the report prepared by the Working Group on the Use of Demographic Analysis in Census 2000. If you would like to convene the working group to discuss the recommendations with the Sampling and Estimation Review Committee, please contact Charlene Leggieri at X3970.

Distribution
DA Working Group Members
- B. Bell (SRD)
- C. Gibson (POP)
- P. Das Gupta
- G. Spencer
- G. Robinson
- M. Mulry (DSSD)
- A. Vacca
- R. Fay (DIR)
- C. Leggieri (DMD)

J. Thompson (DMD)
E. Wagner

TABLE OF CONTENTS

REPORT OF THE WORKING GROUP ON THE USE OF DEMOGRAPHIC ANALYSIS IN CENSUS 2000

1. Introduction

In October 1995, the Sampling and Estimation Review Committee (SERC) established a working group to identify the set of questions that must be answered in order for the Census Bureau to decide if and how estimates derived from Demographic Analysis (DA) could be integrated in the production of final estimates for Census 2000. Further, the SERC asked the working group to identify the necessary research to answer these questions along with recommendations regarding priorities and possible candidates for conducting the research. This report contains our findings and recommendations. Some background is given in Section 2, followed by our recommendations in Section 3. The remainder of the report gives more detail on policy issues and potential research projects that were considered.

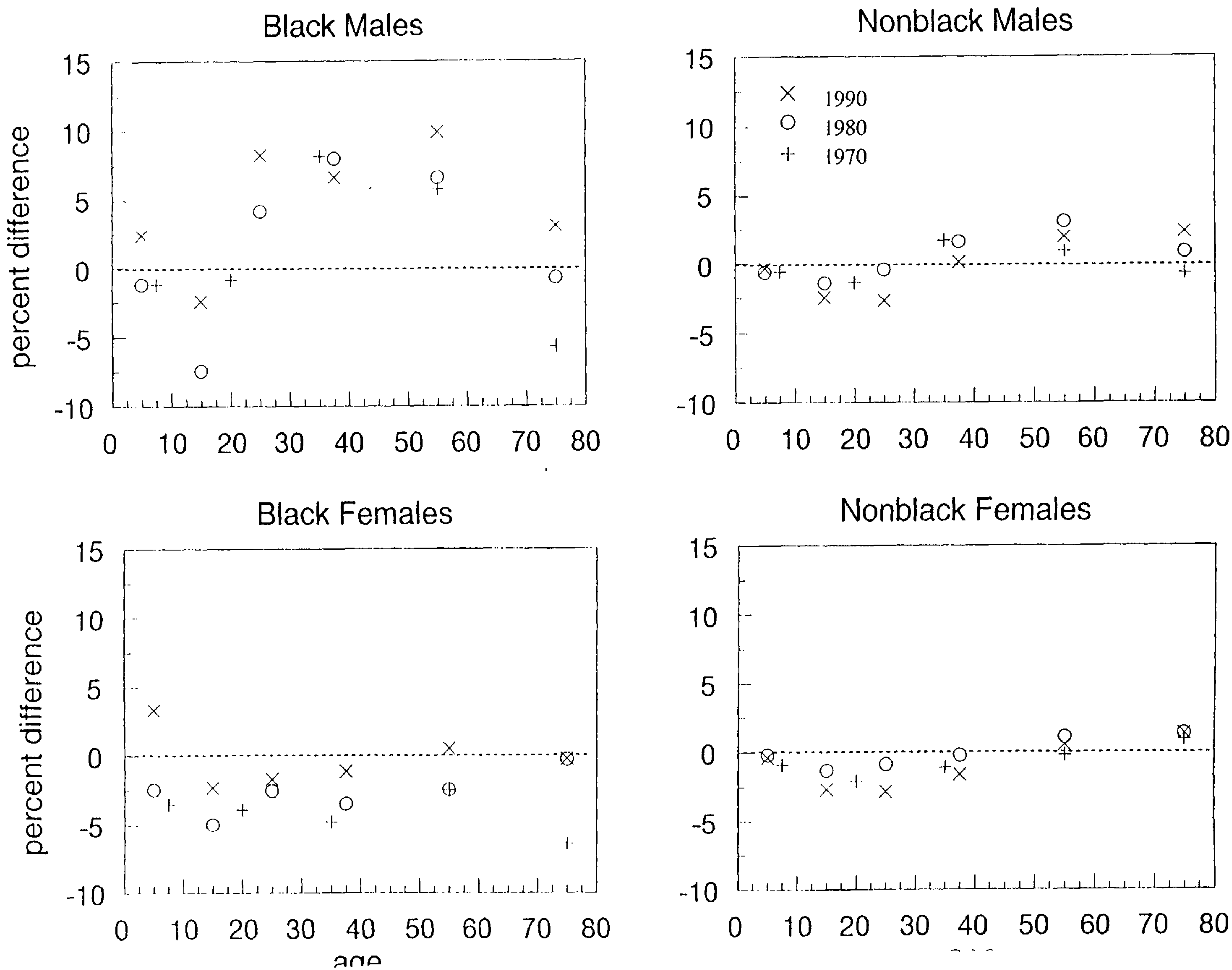
The working group was comprised of representatives from the Decennial Statistical Studies Division (Mary Mulry, Ann Vacca), Population Division (Prithwis Das Gupta, Campbell Gibson, Gregg Robinson, Greg Spencer), Statistical Research Division (William Bell), and the Office of the Director (Robert Fay). Decennial Management Division (Charlene Leggieri) facilitated/chaired the group and Ruth Ann Killion provided direction and guidance for the group.

2. Background

2.1 Motivation -- Differences Between DA and Coverage Measurement Survey (CMS) Results

DA has been used to evaluate census results for many years, as have coverage measurement surveys (CMSs) such as the 1990 Post Enumeration Survey (PES) and 1980 Post Enumeration Program (PEP). (For simplicity, we use the term "coverage measurement survey" as a generic reference to any match study estimating census coverage.) DA can also be used to evaluate corresponding CMS results. Figure 1 shows comparisons between DA and CMS results by age-race-sex for 1970, 1980, and 1990. The results are presented as percentage differences, defined as 100(DA - CMS)/DA, where DA is the demographic analysis population estimate for a given age-race-sex group, and CMS is the corresponding CMS estimate. Values above zero indicate CMS undercoverage relative to DA; values below zero indicate CMS overcoverage. Thus, these percent differences can be thought of as CMS undercoverage rates analogous to the familiar census undercount rates. (Notes: The 1980 CMS is the PEP 3-8 series of estimates, and the 1980 DA results assume 3 million undocumented immigrants. The 1970 CMS was a match study that used unweighted data

Figure 1. CMS and DA Percent Differences for Three Census Years
100(DA - CMS)/DA



from the March 1970 CPS to estimate census omission rates. The 1970 results overstate CMS coverage since there was no estimation made of census erroneous inclusions. The 1980 and 1970 DA estimates represent revised estimates that were made during the production of the 1990 DA estimates--see discussion in Section 5).

The most noteworthy aspect of Figure 1 is the large undercoverage of adult black males in all three CMSs. Differences between DA and CMS for the other age-race-sex groups are less important. There are some estimates of CMS overcoverage (though those for 1970 are overstated); these are rather consistent for ages 10-30, and are particularly large for blacks 10-19 in 1980. There is some evidence that CMSs tend to undercover nonblack males above age 30, but there is little evidence for persistent, significant undercoverage of black or nonblack females.

Differences between DA and CMS estimates can arise from errors in either. It is generally assumed, however, that the persistent differences between DA and CMS results for adult black males are due to underestimation in the CMS. There are two general reasons for assuming this. The first is that similar differences do not occur for black females. Errors in DA would not be expected to be so greatly different for males and females as to lead to the large differences observed for adult black males but not for adult black females. The second reason is that a reasonable explanation for underestimation of males in the CMS is *correlation bias*, a violation of the independence assumption that underlies the DSEs. It can occur either because (1) different individuals within poststrata have different probabilities of being included in the census and/or the CMS (heterogeneity), or (2) the act of being included in the census tends to make it more likely that someone will also be included in the CMS. The results in Figure 1 suggest correlation bias is present in CMS estimates of adult black males, may be present in CMS estimates of nonblack males over 30, and probably is not present in CMS estimates of females.

Much of the following discussion on combining DA and CMS results refers to combining DA and DSEs, rather than referring more generally to combining DA with other potential CMS (ICM) estimators (for example, Census Plus). This is for convenience of exposition and because previous research has focused on combining DA and DSEs. Section 6 considers a research project on combining DA and Census Plus or even census post-NRFU estimates. It should be noted, however, that previous research using data from the 1980 and 1990 censuses, and the 1988 and 1995 test censuses, has shown simulated Census Plus estimates to be significantly lower than corresponding DSEs, so that differences between DA and Census Plus for adult males are even larger than those between DA and DSE results. Corresponding differences between DA and census post-NRFU estimates would be larger still.

2.2 Historical Background

Given the persistent difference observed between DA and CMS results for adult black males, it is natural to ask whether these results can be combined to remove the correlation bias suspected in the CMS estimates. This would be consistent with one of the major goals of Census 2000, that of reducing differential undercount. Research has been done at the Census Bureau on methods of combining DA and CMS results, and such methods were considered, though ultimately rejected, for producing the official 1990 PES estimates.

Kirk Wolter did research on the combining issue in the 1980s, ultimately publishing a paper on the subject (Wolter 1990). He suggested sex ratios from DA could be used as an additional piece of information to estimate the "cross product ratio" (a measure of dependence in the 2x2 census-CMS table), and hence produce combined DA-CMS estimates. His approach applied only at the national level within age-race-sex groups (the level of the DA data to be used), leaving open the question of how to produce subnational combined estimates. In work for the 1990 PES, William Bell generalized Wolter's approach to address this issue. Without going into details (for which see Bell (1993)), Bell's approach (1) uses the usual DSEs for females (assumes independence), (2) selects a model for males that produces alternative poststratum DSEs allowing for some dependence between the census and CMS, and (3) estimates the dependence by controlling the alternative male DSEs to reproduce the DA sex ratios when aggregated to the national level.

Other researchers at the Census Bureau who have explored different but related approaches to combining include Isaki and Schultz (1986) and Das Gupta and Robinson (1990).

The original research plans for the 1990 undercount adjustment included combining PES estimates with DA sex ratios. This issue was revisited, however, following the Commerce Department's initial decision in 1987 not to adjust and the subsequent court settlement (which stipulated that a decision on census adjustment be made in July, 1991). William Bell and Howard Hogan then made presentations on the combining methodology to the Undercount Steering Committee (USC) for their review of technical methods considered for the 1990 PES. They also made a presentation to the Commerce Secretary's Advisory Panel. Charles D. Jones, Associate Director for the Decennial Census, ultimately made the decision not to combine, a decision supported by the USC.

In "Decision on Combining PES and DA Estimates," a June 3, 1991 Memorandum for the Record (included as Attachment 1), Jones documented eight reasons for the decision not to combine. While we will not reconsider these reasons in detail here, we feel a few general comments are warranted. (1) Some of the reasons cited were particular to the 1990 census and would not apply to Census 2000. (2) Other reasons cited, on the other hand, still pose legitimate objections to combining. Particularly important are those reasons expressing concern about errors in the DA estimates or about the validity of assumptions underlying any combining method. The latter issue is discussed in the next section, and the former is reflected in the discussion of research projects to improve DA in Section 5. (3) The two reasons reported as weighing heaviest in the decision not to combine stated some additional concerns about the complexity of the combining procedure proposed. We would suggest that this objection be reconsidered for Census 2000, and would point out that a combining method need not be complicated. In fact several proposed approaches yield combining methods analogous to standard raking procedures routinely used to enforce population control totals in many surveys.

2.3 How Important is Uncertainty About Combining Methods: A Major Issue on Which There is not a Working Group Consensus

Bell (1993) showed that alternative methods of combining are available that control to the same national DA information, but that produce different subnational estimates. This means that, given just the census, CMS, and DA data, there is some uncertainty about how to combine that translates into uncertainty about subnational estimates. How much of this uncertainty there is depends on how wide a range of combining methods one considers reasonable. One of the research projects recommended in Section 3 is to study further how much variation there is across "reasonable" combining methods, taking into account how much uncertainty there is in basic CMS estimates (from sampling error and methodological decisions such as choice of poststratification).

Despite lengthy discussions, strong differences of opinion remained among working group members. Given this situation, the working group decided that it would be best to present two differing views to reflect the range of thinking on this issue. Appendix A (Different Ways to Combine Yield Different Subnational Results), written by Bell, emphasizes that the CMS and DA data provide no basis for discriminating among alternative combining methods -- all are equally good as far as these data are concerned though producing different subnational estimates. Appendix B (Demographically Meaningful Ways to Combine Yield Similar Subnational Results), written by Das Gupta, contends that sufficient demographically meaningful assumptions can be made to significantly restrict consideration to a set of combining methods that will yield similar subnational results. Clearly, the issue of uncertainty about combining methods will be a major factor in the Census Bureau's decision on whether or not to combine.

2.4 Assumptions Made by the Working Group

In developing the set of questions and research projects, the working group made the following assumptions:

- We assumed that DA will be used as an evaluation tool in 2000, as it was in 1990, even if the decision is made not to integrate the DA estimates with ICM results. Therefore, at least the same level of work that went into developing the national DA estimates for 1990 is assumed for 2000.

- We assumed that subnational DA estimates would not be used for combining, but may have other uses noted below and as described in Appendix F.

- We assumed that we cannot use site tests or a dress rehearsal to test the efficacy of combining with DA because the "standard" DA estimates are at the national level. New subnational DA indicators and analytic techniques (sex ratio analysis) can be useful ICM evaluation tools for the dress rehearsal, however.

- We assumed that combining with DA could be considered with either estimation method chosen for ICM (CensusPlus or DSE).

- We assumed that our primary focus was not to lay out research to support use of DA to adjust the census numbers without a coverage measurement survey (e.g., if ICM fails or is late).

- We did not allow limitation of resources to constrain our thinking about what research is necessary or desirable. Some constraints must follow, of course, given a real world environment of limited resources.

- We assumed the following definition of cost levels to carry out any of the research projects considered:

  Low cost projects need less than one FTE for one year.
  Medium cost projects need about one FTE for one year.
  High cost projects need more than one FTE for one year and/or field work.

- We assumed that the only racial breakdown available from DA estimates will be black/nonblack. Further, we assumed that there would be no major change in the census race question for 2000. (If there IS a major change in the race question, then the usefulness of a black/nonblack breakout of the DA estimates would have to be reconsidered.)

3. Recommendations

The working group discussed various research projects that would be desirable for answering questions about if and how DA and ICM results should be combined for Census 2000. Projects were assigned low, medium, or high priority, and assessed as low, medium, or high cost. This section gives our recommendations of the most important research projects to pursue. Those projects that were assigned high priority and low cost form the core of our recommendations. The projects are drawn from three general areas: research to improve DA, research on combining methodology, and research on statistical estimation. Sections 5, 6, and 7 discuss all the projects considered in these areas, including those not identified as most important in this section.

Some of the research projects we recommend overlap with projects being independently contemplated in other contexts. For example, almost all of the research projects on statistical estimation discussed in Section 7 were drawn from the document, "Research Topics for Studying Sampling and Estimation Methodology for Census-Taking," (September 19, 1995) by Mulry, Singh, Woltman, and Robinson. In fact, many of the projects we recommend are desirable for other purposes, and thus may go forward regardless of the decision made about combining. We did not, however, attempt to take into account any such side benefits of various projects when assigning priorities.

3.1 Recommended Research to Improve Demographic Analysis

Potential research that can improve the demographic analysis (DA) estimates is discussed in Section 5. We identify eight project topics where research could lead to improvements in the DA estimates (see Table 2). When prioritized in terms of impact and cost, the four most important research projects are:

1. Undocumented immigration
2. Emigration
3. Research on race inconsistencies
4. Uncertainty intervals

Research on undocumented immigration and emigration is needed to make methodological improvements and to maintain the quality of the DA estimates (unlike births, deaths, and legal emigration, current administrative data on undocumented immigrants and emigrants are not available). Research on race classification is needed to assess the degree of inconsistency in the race categories of the DA estimates and the census. Research is needed to improve the statistical assessment of uncertainty in the DA estimates.

3.2 Recommended Research on Combining Methodology

The most important research projects on combining methodology are activities 1 and 6 in Section 6. Activity 1 seeks to answer the question, "What information (at the national level) from DA should be used in combining?" The approach considered for 1990 used sex ratios by black/nonblack and the PES age groups. However, use of other information from DA should be investigated, particularly use of DA age distributions in addition to the sex ratios.

Activity 6 investigates the range of variation across reasonable alternative methods of combining. If this range is judged to be large, then there is considerable uncertainty about how to combine. If it is judged to be small, then there is little uncertainty about how to combine. These judgements must take into account the range of uncertainty in the basic CMS estimates.

Two other important projects are activities 2 and 3 in Section 6. Activity 2 addresses the question of how combining can be done in a way that produces as an end result a file with all persons assigned to households or group quarters. This activity overlaps with research on doing this for basic ICM estimation. Activity 3 addresses the question of whether combining should try to estimate additions and deletions to "special populations" such as American Indians on reservations, certain group quarters populations, etc. This issue was avoided in 1990 by modifying DA results before combining to cover only the civilian noninstitutionalized population estimated by the 1990 PES.

3.3 Recommended Research on Statistical Estimation

The most important statistical estimation research relevant to the combining issue is covered under activities 1.a, b, and c in Section 7. These all involve studying whether modifications to the DSE approach used in 1990 can produce estimates closer to DA, thus reducing the correlation bias problem that motivates consideration of combining in the first place. Activity 1.a examines if alternative poststratifications of the 1990 PES data (e.g., using the targeting database) can accomplish this goal. Activity 1.b explores extensions of the logistic regression approach of Alho, Mulry, Wurdeman, and Kim (1993) to reducing correlation bias. Finally, Activity 1.c examines extensions to an approach called split DSE that was developed by William Bell. Although activity 1.c is assigned only medium priority in Section 7, it is recommended here because it should be extremely low cost, i.e., the extensions being considered should require very little effort to investigate.

The next most important research is activity 2, which investigates the occurrence of negative cells in poststratum 2x2 tables in the 1990 PES. Negative cells occurred when the PES estimated more matches than the census marginal total for a poststratum. Since the matches were a sample weighted estimate, while the census marginal total used was the count less a sample weighted estimate of erroneous enumerations, sampling error alone could produce negative cells. However, the large number of occurrences of negative cells in 1990 (about 1/3 of the poststratum 2x2 tables had negative cells) suggests that other errors (e.g., geocoding errors) may have contributed to this problem. Research that helps us understand and possibly correct problems leading to negative cells would provide a firmer foundation for combining methods such as those of Bell (1993) which make use of the cells of the 2x2 tables.

4. Policy Issues

Question: What is the goal(s) of the "one-number" census?

Background: The May 19, 1995 document describing the Reengineered 2000 Census states (p.III-1) that, "the primary purpose of the 2000 census is to provide a complete national enumeration, together with comprehensive data that describe the characteristics of the population for each geographic entity." While this document clearly articulates the desired product resulting from the 2000 Census, the working group believes that one of the factors in making a decision about incorporating estimates from Demographic Analysis into the "one-number" census, is an explicit understanding of the goal or goals of such a census. To stimulate discussion on this issue, we have listed below some potential goals of the one-number census. The discussion may determine that some of these are not goals or that some are more important than others, but the answer to the policy question posed will provide guidance in answering the question about integrating DA into the census results.

Potential goals of a one-number census:

To provide the most accurate decennial census estimates possible for meeting legal requirements for congressional apportionment and redistricting. The goal of the one-number census is to provide accurate data to meet the legal requirements for Congressional apportionment and redistricting. The size of the population for states meets the requirement for apportionment. The requirement for redistricting is the number of people in each race/Hispanic ethnicity category for the population 18 years of age and over in each block. Not placing additional requirements for accuracy of other characteristics is in keeping with the strategy to keep the methodology for determining census numbers as simple as possible. When more than one measurement method is used in the creation of the census numbers, the Census Bureau must explain all the measurement methods. Also, the Census Bureau has to defend assumptions underlying the methodology for incorporating the estimates from other measurement methods in the census numbers. By focusing only on the representation requirements, this approach avoids complicating the census process any more than necessary.

To prepare decennial census estimates of the population using a methodology that is as simple and understandable as possible, while still addressing issues of cost and differential undercount. Because the use of estimation is a departure from traditional counting methods, we need to develop the estimates in ways that are understandable and therefore credible to the public. Certainly it is not reasonable that everyone will understand technically sophisticated assumptions and algorithms that go into the "one-number" census, but the overall strategy for measuring (estimating) the completeness of the counting effort and then correcting for errors should be (intuitively) understood. One of the factors in the decision not to combine DA and PES estimates in 1990 was the concern that the procedure would not be understood by even knowledgeable undercount experts (see Attachment 1).

To provide the most accurate decennial census estimates possible of the total population by race, Hispanic origin, sex, and single year of age. This goal reflects the general desirability of improving the accuracy of characteristics data where possible. See the following goal for a specific example.

To provide estimates that are satisfactory as benchmarks for the postcensal population estimates and projections programs. The 2000 census data on race, Hispanic origin, sex, and age will be used in the Bureau's postcensal population estimates and projections program, which includes providing population controls for national demographic surveys, including, most notably, the Current Population Survey and the Survey of Income and Program Participation, and in the American Community Survey. At present, because of policy decisions that are not demographically consistent, the various census benchmarks are not consistent. The Bureau's population estimates and projections are census enumeration level consistent. The controls for demographic surveys are consistent with the 1990 census population adjusted for net census undercount based on the PES. In addition to being inconsistent with the basic estimates and projections, these controls yield implausible patterns of population by sex and single year of age compared to DA estimates because the 1990 PES did not capture past demographic patterns reflecting annual birth registration data and sex ratios by age, due presumably to the limitations of sample size and to correlation bias in the PES.

5. Research to Improve Demographic Analysis

Question: How can the DA estimates be improved?

Background: National DA estimates of the population under 65 in 2000 are based on historical data on the components of change: births, deaths, legal immigration, undocumented immigration, and emigration. The DA estimates for the population 65+ are based on Medicare data. With the exception of undocumented immigration and emigration, the component estimates are derived from administrative data that cover essentially the entire population. No sampling is involved for the core components of births, deaths, and legal immigration (though some sampling is involved in developing correction factors for birth underregistration).

The procedures for producing DA estimates for 2000 are largely in place now. Historical component data were compiled and analyzed in the production of the 1990 DA estimates. Using current administrative sources (births, deaths, Immigration and Naturalization Service (INS), statistics on legal immigrants) and estimates of undocumented immigration and emigration, the estimates have been updated monthly since 1990 as part of the Census Bureau's ongoing population estimates program. Table 1 shows the components that comprise the overall DA projections for 2000.

While the national DA estimation procedures are well developed, continuous research is needed to maintain or improve the quality of the estimates. This would continue the Census Bureau's demographic research program of the past 35 years, which has produced improvements in data, assumptions, and methodology. Demographic research intensified during the 1990 census cycle, leading to (1) the first-time development of uncertainty models for the DA estimates (documented in 1990 Demographic Evaluation Project D11), and (2) detailed assessments of the estimates of individual components used to construct the DA estimates (documented in 1990 Evaluation Projects D1 to D10).

TABLE 1
PROJECTIONS OF COMPONENTS OF DEMOGRAPHIC ANALYSIS FOR 2000

Component	Number (in millions)	Percent
TOTAL	277	100
Births (1935-2000)	232	84
Deaths (1935-2000)	-14	-5
Legal immigration (1935-2000)	24	9
Undocumented immigration (1935-2000)	5	2
Emigration (1935-2000)	-5	-2
Medicare (in 2000)	35	13

Research Activities: Building on the foundation set by the 1990 DA program, a comprehensive demographic research program for 2000 would include two major activities: (1) to re-evaluate and update the assumptions that underlie the component estimates at the national level (e.g., assumptions about birth and death registration completeness) and (2) to improve where needed the methodology and data input of the component estimates, drawing on new data sources where possible (e.g., undocumented immigration and emigration). Table 2 lists the specific research activities needed to maintain and improve the national DA estimates. The projects are prioritized (high, medium, low) on different dimension, and relative costs are indicated. Appendix C provides more detail on the individual research projects; Appendix D identifies the criteria used to assign priorities.

It is important to note that the continuous research on the underlying DA components--which leads to changes in assumptions and new data sources--consequently changes the initial DA population estimate for each census (e.g. the original DA estimates from the 1970 and 1980 censuses have changed, as will the estimate for 1990 and probably 2000). The effect of these revisions was studied in Demographic Analysis Evaluation Project D10 and some subsequent work. This research demonstrates that DA revisions have a much lesser effect on DA proportionate distributions than they do on DA levels. For example, as shown in Table 3 and illustrated in Figure 2, the DA sex ratios for Blacks vary little across the different revisions--and all are significantly greater than the census sex ratios. We can attach greater confidence to the age-sex proportionate distributions and sex ratios derived from demographic analysis than in the "point" estimates for any given group, and it is this distributional attribute of the DA estimates on which our research on combining focuses.

Table 2. DA Research Priorities and Relative Costs
(See attachment for criteria used to assign priorities)

Research Projects	Priority Category						Cost (In-house projects)
	Impact on DA total population	Impact on age-sex structure	Impact on race structure (Black/ Non-Black)	Impact on variance measurement	Need because lack of current data	Overall impact	
<u>Basic components</u>							
1. Births							
Basic research	H	L	M	M	L	M	M
BRC Test	L	L	L	L	H	M/L	H
2. Deaths	M	L	L	L	L	L	L
3. Legal immigration	M	L	L	M	M	M/L	L
4. Undocumented immigration	L	H	H	H	H	H	M
5. Emigration	L	M	M	H	H	H/M	M
6. Medicare	M	L	L	M	L	M/L	M
<u>Other research</u>							
7. Race Inconsistencies							
Basic research	L	L	H	M	H	M	M
Longitudinal match	L	L	H	M	H	M	H
8. Uncertainty intervals	L	L	L	H	M	M	L

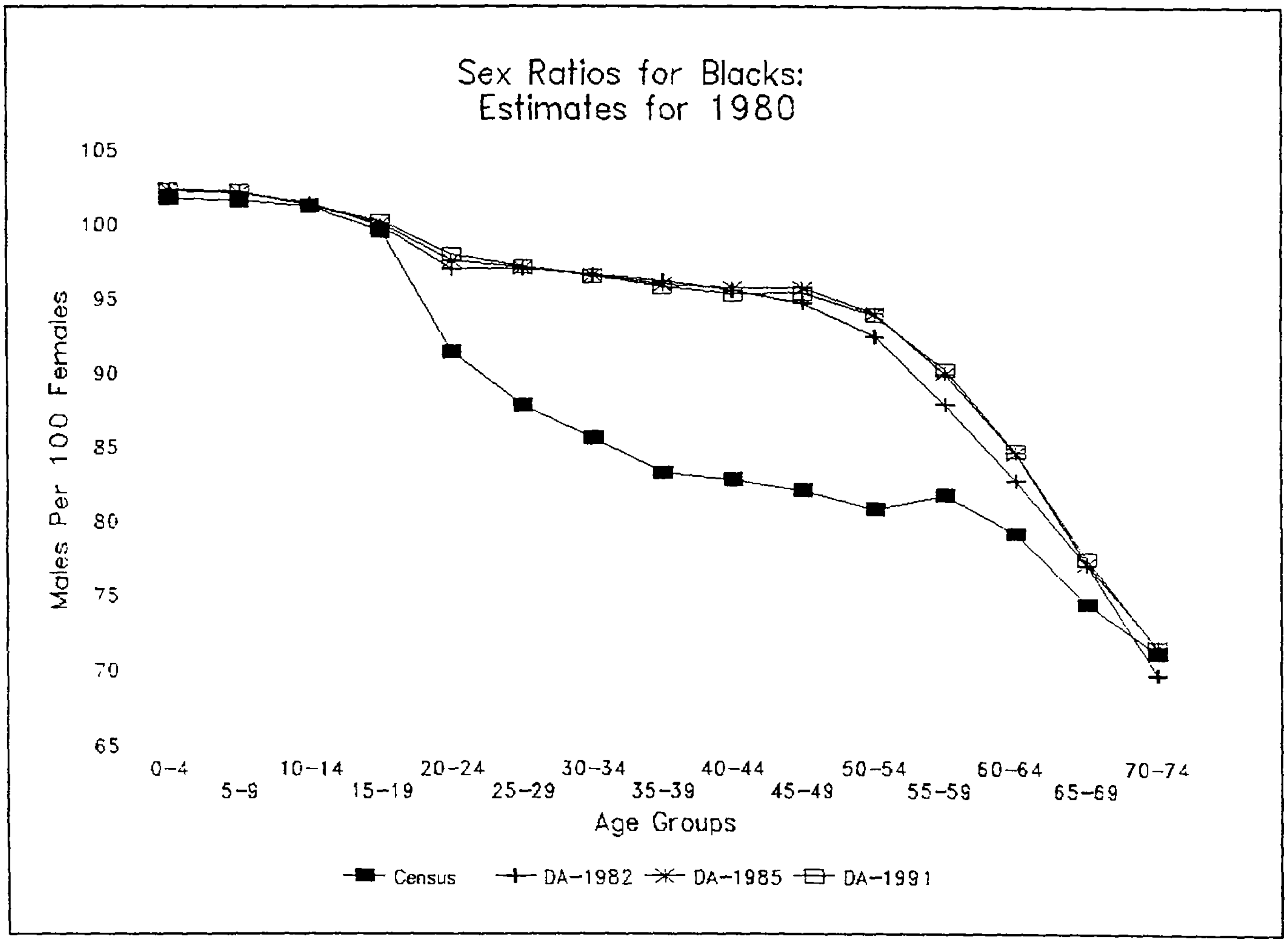
TABLE 3

Comparison of "Initial" and Subsequent Revisions to Demographic Analysis Estimates of Sex Ratios and Percent Net Undercount: 1980 Census

	Sex ratios				Percent Net Undercount					
					DA 1982		DA 1985		DA 1991	
	Census	DA 1982	DA 1985	DA 1991	Male	Female	Male	Female	Male	Female
All ages	89.6	94.8	95.2	95 3	7.5	2.1	8.8	3.1	7.5	1.7
0-4	101.6	102.2	102.3	102.3	9.2	8.7	9.6	9.0	8.9	8.3
5-9	101.6	102.1	102.2	102.2	5.9	5.5	6.1	5.6	5.7	5.1
10-14	101.2	101.4	101.3	101.2	1.1	0.9	1.8	1.7	1.4	1.3
15-19	99.6	99.9	100.1	100.2	-0.8	-1.0	0.3	-0.2	0.3	-0.4
20-24	91.4	97.0	97.6	97.9	7.2	1.6	8.9	2.8	9.2	2.8
25-29	87.8	97.0	97.1	97.2	11.2	1.9	12.7	3.5	12.7	3.3
30-34	85.7	96 6	96.6	96.5	12.0	0.8	13.4	2.4	12.4	1.4
35-39	83.3	96.2	96.0	95.8	16.8	3.8	16.6	3.9	13 7	0.8
40-44	82.8	95.5	95.7	95.3	15.4	2.4	18.5	5.8	13.2	0.0
45-49	82.1	94.7	95.7	95.4	16.7	3.9	18.4	4.9	13.6	-0.4
50-54	80.8	92.5	94.0	93.9	13.6	1.1	15.7	1.9	11.4	-3.0
55-59	81.8	87.9	90.0	90.2	7.8	1.0	10.8	1.8	7.1	-2.5
60-64	79.2	82.8	84.6	84.8	4 1	-0.2	6.7	0.4	5.1	-1.5
65+	68.3	67.4	67.5	67.9	-2.7	-1.5	-1.4	-0.3	-1.8	-1.2

Note: DA-1982 refers to initial demographic analysis estimates (produced in 1982); DA-1985 refers to "final" set in the 1980 cycle (produced in 1985); DA-1991 refers to revised estimates for 1980 produced as consequence of development of 1990 estimates (1990 census cycle)

Figure 2


Sex Ratios for Blacks:
Estimates for 1980
Males Per 100 Females
105
100
95
90
85
80
75
70
65
0-4
5-9
10-14
15-19
20-24
25-29
30-34
35-39
40-44
45-49
50-54
55-59
60-64
65-69
70-74
Age Groups
Census
DA-1982
DA-1985
DA-1991

Question: How can the uncertainty measures of DA estimates be improved?

Background: At the national level, Demographic Analysis (DA) data provide population by single years of age, sex, and race (Black/Non-Black). Extensive research has been done on uncertainties in DA levels over the past several years, and an exhaustive description of the research is provided in Demographic Analysis Evaluation Project D11.

Initially, several simulation models were attempted to study uncertainties in DA estimates. Basically, these models involved the assumptions of two certainty limits for each component, independence of the components, and a distribution between the certainty limits of a component. A set of random values were drawn from these distributions to form one DA estimate. A very large number of such estimates (say, 10,000), when arranged in ascending order, demarcated a 95-percent or a 99-percent error interval. The assumption of independence of the components was later replaced by some constraints on the choice of the random values drawn from the distributions. The certainty limits had been arrived at by a judgmental consensus of the Census Bureau experts knowledgeable about estimation methodology and possible errors in the components of change. These simulation models were presented at the Census Bureau Advisory Committee Meetings in October 1987, and based on the Committee's advice, we switched from simulation to analytical techniques to study uncertainties in DA.

The general approach in the analytical models is that if the means and variances of the individual DA components and the correlations between them are somehow estimated, then the mean and variance of the DA estimate, which is the sum of the DA components, can be estimated and hence an error interval can be constructed. In order to compute the mean and variance of a component, this approach needs a probability interval (in terms of high and low multipliers) around each component and the estimation of this interval is the most crucial part in the analytical models. Ten Demographic Analysis Evaluation Projects (D1-D10) were carried out by staff members to address specific sources of uncertainty in individual components. These studies have been very instrumental in our effort to combine the various possible sources of error in a component in the form of an error interval in terms of two multipliers.

Various analytical models were developed depending on whether the point estimate of a component was treated as the mean, median, or mode of the distribution and also on whether the distribution of the component was normal, gamma, or something else. These models were thoroughly discussed in the Census Bureau Technical Design and Estimation Committee meetings. This Committee finally approved the model which assumes a normal distribution for a component with a mean equal to the average of the high and low multipliers (the variance being determined from the normal deviate associated with the probability limits). The analytical models were presented at the Census Bureau Advisory Committee Meetings in

October 1989. The Committee members also approved the same model with normal distribution because it was the simplest to interpret and easily understood.

Research Activities: If a decision is made to integrate DA with the Census/Survey estimate for the One-Number Census in 2000, the seriousness of the problem of uncertainties in the DA estimates would depend to a great extent on whether we use the DA levels or the DA age-sex proportionate distributions in the models. Obviously, the use of DA levels would introduce larger errors, whereas the problem of uncertainties would be much less serious when the DA age-sex proportionate distributions are used. Since the DA age-sex proportionate distributions are derived from the DA levels, the research on uncertainties in DA levels is also relevant to the study of uncertainties in DA age-sex proportionate distributions. As mentioned before, we have dealt with various techniques for studying uncertainties in estimated DA levels. An inter-divisional team should be formed to review these techniques to assess their relevance to the study of uncertainties in the DA age-sex proportionate distributions. Keeping in view the earlier research, the team will attempt to produce some uncertainty measures that will be acceptable to the statistical community. For an alternative approach to dealing with uncertainty in DA age-sex proportionate distributions directly and not through DA levels, see Appendix E.

6. Research on Combining Methodology

Question: What method(s) should be used to combine DA and ICM estimates?

Background: As discussed in Section 2.2, research done prior to the 1990 census developed methods of combining DA and 1990 PES results. Additional research could seek to extend this work, or to explore alternative methods of combining.

Research Activities:

1. Investigate what information from DA (at the national level) should be used in combining. The focus of the 1990 research was on use of DA sex ratios by age-race (black/nonblack) groups, although use of DA population totals by age-race-sex groups was also considered. Another possibility may be to use both sex ratios and age distributions, to maintain, to the extent possible, the age-sex structure of DA. (Use of male and female pop totals by age-race groups does maintain the age-sex structure, but use of sex ratios by age-race groups does not, since only the estimates for males are changed.) Also, can combining be done to maintain approximately the single year-of-age structure of DA?
Priority: High Cost: Low

2. Study how ICM estimates can be combined with DA results in a way that produces an extrapolated population file, i.e., a file with all persons assigned to households (or group quarters -- see 3.). (Note: This overlaps with research on how to do this for basic ICM estimation.)
Priority: High Cost: Medium

3. Study if and how additions or deletions due to combining should also be allocated to "special populations" (American Indians on reservations, armed forces, homeless, certain group quarters populations, etc.). A sensitivity analysis under alternative assumptions could be performed. (Note: This issue may not be a problem if ICM estimation is **not** required to assign persons to households or group quarters -- see 2.)
Priority: Medium Cost: Low

4. Study differences in results of combining Census Plus or even census post-NRFU estimates with DA from those of combining DSEs with DA.
Priority: Low Cost: Low

5. Examine application of O'Connell's generalization of Bell's (1993) approach to combining.
Priority: Medium Cost: Medium

Question: What is the range of variation across "reasonable" combining variants, and how does this compare to the range of uncertainty in basic CMS estimates?

Background: As discussed in Section 2.3 and Appendix A, alternative methods of combining are available that control to the same national DA information, but that produce different subnational estimates. If the range of variation among reasonable combining alternatives is judged to be "large," this reflects considerable uncertainty about how to combine. If the range of variation is judged to be "small," it reflects little uncertainty about how to combine. This judgement must take into account the range of uncertainty in the basic CMS estimates.

Research Activity:

6. Explore a range of alternative combined estimates for 1990 PES data, drawing on work of Bell (1993), Das Gupta and Robinson (1990), and material presented in Appendix A. Investigate the range of "reasonable" alternatives, making judgements about which alternatives are reasonable and which are not. Compare this range to the analogous range of uncertainty about "reasonable" 1990 PES estimators based on comparable data (e.g., produce 1,392 poststrata results using data from the reworked matching for comparison to the 357 poststrata results). Also take into consideration sampling error in the 1990 PES estimates. Comparisons should evaluate the ranges for population shares, not just totals.

7. Research on Statistical Estimation

Question: What research problems on statistical estimation of census coverage are there that, if addressed, would help make the decisions on whether and how to combine, or could improve combined estimates?

Background: Previous research has pointed out certain problems in CMS estimation of census coverage (correlation bias, level of sampling variability, limitations of synthetic assumptions, etc.) Improved ICM estimation techniques should lead to improved combined estimates. Also, if it were possible to improve CMS estimates to essentially eliminate correlation bias, this would remove the major motivation for combining. (Note: Some of the research activities listed below may already be planned as part of ICM estimation research. Their importance to this general activity may be different than the priorities assigned here, which reflect their importance to the combining issue.)

Research Activities:

1. Study alternative ICM estimators with potential to reduce correlation bias, which could reduce or eliminate the need to combine. A major motivation for combining is to address the "correlation bias" problem believed present in DSEs, particularly for adult black males, in the 1990 PES and previous census coverage measurement surveys. The research could examine the following variations on DSE:
   a. Investigate alternative poststratifications determined from the targeting database or otherwise using 1990 data.
   Priority: High Cost: Low
   b. Model heterogeneity in capture probabilities -- extend work of Alho, Mulry, Wurdeman, and Kim (1993).
   Priority: High Cost: Low
   c. Further investigate split DSE (Bell) and its extensions.
   Priority: Medium Cost: Low
   d. Investigate triple system estimation using administrative records or using two surveys plus the census. (Note: Feasibility is a concern here both in regard to doing two surveys and in regard to whether administrative records data for the entire country can be cleaned up (i.e., remove erroneous inclusions) to the point where it is suitable for triple system estimation.)
   Priority: Low Cost: High
   e. Could other uses of administrative records data improve ICM estimates and bring them closer to DA results? Using the 1996 Community Census (ICM Test), evaluate the extent to which use of administrative records to enhance ICM reduces bias.
   Priority: Medium Cost: Low

2. Some of the methods investigated by Bell (1993) for combining involved distributing the increase in the number of males proportional to cells or sums of cells

of the 2x2 tables for poststrata. About 1/3 of the poststratum 2x2 tables in the 1990 PES had negative cells (more matches estimated than correct census enumerations). Research could investigate what caused this, and how it may be avoided through changes in operation or in estimation (including use of "smoothing" in determining the 2x2 table entries rather than smoothing the adjustment factors).
Priority: High Cost: Medium

3. The combining approach investigated by Bell (1993), and most Census Bureau work on ICM estimation for that matter, provide "adjustment factors" by poststrata. These are then synthetically carried down to produce adjusted estimates by demographic groups at the block level (followed by controlled rounding). Research could consider alternatives to the synthetic assumption and their effects on the results of combining DA with ICM estimates. 1.b above is one possible approach. Another could involve relating the block adjustment factors to block characteristics available from other (e.g., administrative records) data.
Priority: Medium Cost: Low/Medium

4. Study if ICM estimates can be produced for single year-of-age groups, to facilitate combining to maintain the DA single year-of-age structure (see number 5). Also consider if the large planned ICM sample will support direct estimates for single year-of-age (or slightly less detailed age groups), that would be adequate for use in combining if not for use as stand-alone estimates.
Priority: Medium Cost: Medium

5. Study if "smoothing" ICM estimates, as done for the 1990 PES, has potential to improve combined estimates. This includes studying if combining should be done before or after the smoothing. (Note 2. above.)
Priority: Medium Cost: High

6. Study how variances of combined ICM/DA estimates should be produced. A particularly difficult question is how to produce variances of combined estimates for small areas? (Note: Research on variance estimation would depend on the nature of the ICM estimates and combining methodology to be used, so most of the work on variances should wait until these things are determined.)
Priority: Medium/High* Cost: Medium

*Priority is high for variances of the most important estimates.

8. References

Alho, Juha M., Mulry, Mary H., Wurdeman, Kent, and Kim, Jay (1993) "Estimating Heterogeneity in the Probabilities of Enumeration for Dual-System Estimation," Journal of the American Statistical Association, 88, 1130-1136.

Bell, William R. (1993) "Using Information From Demographic Analysis in Post-Enumeration Survey Estimation," Journal of the American Statistical Association, 88, 1106-1118.

Das Gupta, Prithwis and Robinson, J. Gregory (1990) "Combining Demographic Analysis and Post-Enumeration SUrvey to Estimate Census Undercount," unpublished manuscript, Population Division, Bureau of the Census.

Isaki, Cary T. and Schultz, L.K. (1986) "Dual-System Estimation Using Demographic Analysis Data," Journal of Official Statistics, 2, 169-179.

Mulry, Mary H., Singh, Rajendra P., Woltman, Henry and Robinson, J. Gregg (1995) "Research Topics for Studying Sampling and Estimation Methodology for Census Taking", Bureau of the Census, September 19, 1995.

Wolter, Kirk M. (1990) "Capture-Recapture Estimation in the Presence of a Known Sex Ratio," Biometrics, 46, 157-162

ATTACHMENT I

UNITED STATES DEPARTMENT OF COMME
Bureau of the Census
Washington, D.C. 20233

OFFICE OF THE DIRECTOR

JUN 03 1991

MEMORANDUM FOR The Record

From: Charles D. Jones
Associate Director for Decennial Census

Subject: Decision On Combining PES and DA Estimates

Over the past few years, and especially in the recent several months, the technical staff of the Census Bureau concerned with undercount estimation have worked with others to develop a procedure and rationale for combining the PES and demographic estimates of undercount prior to smoothing. The problem is an extremely difficult one because there exists no acceptable subnational estimates of demographic analysis population distributions, and unverifiable assumptions must be made to distribute the estimates to subnational areas. The technical staff expended great effort and ingenuity to try to resolve the problem and are to be commended for their work. They reported the results of their work to a recent Undercount Steering Committee (USC) for review and discussion.

Following that meeting, and after additional consultation, review, and consideration, I have decided that the PES and demographic analysis estimates would not be combined for producing 1990 subnational estimates of census coverage. This decision is supported, in some cases quite strongly, by the vast majority of USC. While a few of the USC were "on the fence" about the decision, there was no strong objection to the decision.

A number of problems were identified with the proposed procedure and rationale including:

1. The preliminary nature of demographic analysis and expectation that these will change dynamically, and perhaps dramatically, over time.

2. Assumptions in demographic analysis about the size of the illegal alien population, emigration, and consistency of racial classification among various data series.

3. Timing; that is, the need to have the estimates by early April in order to use them.

4. Applicability of criteria established to decide whether or not to make the combination.

5. Concerns about the logic of the estimates; that is, the net error in males would be distributed by a ratio of gross nonmatches in a post-stratum to total gross nonmatches in all post-strata (more specifically, the ratio of cells 2,1 + 2,2 in a post-stratum to the sum of cells 2,1 + 2,2 in all post-strata).

6. The concern about whether the data to which the procedure would be applied--all nonmatches--was really the appropriate variable. Some speculated that correlation bias may be more a within household problem and not as much a problem with whole household misses.

7. The lack of knowledge, experience, and understanding by Bureau staff about the validity of the procedures and the underlying assumptions.

8. The concern that by using this complicated and relatively little understood procedure with its assumptions, etc., applied to the model for Dual System Estimation with all its assumptions and deficiencies would yield a result not understood by even the most knowledgeable undercount experts.

All of these problems and concerns weighted into the decision, but the latter two probably were the more persuasive.

P.S. #1: When the demographic analysis data become available, it will be interesting to see whether the data would have passed one of the threshold criteria for combination--namely, Case II in the January 24, 1991 report by William Bell regarding Female Undercount Estimates in the PES and DA.

P.S. #2 In hindsight it appears that the decision not to combine may have been fortuitous. That is, the 3 weeks saved on the schedule was essential to provide time for correcting PES files and for review of smoothing results. Without this additional time the time schedule would have been more seriously threatened than it is now.

APPENDIX A

Different Ways to Combine Yield Different Subnational Results

Bell (1993) compared four different models for producing combined DA-CMS estimates. These four models all were estimated by controlling to national 1990 DA sex ratios for age-race groups, but all produced different subnational estimates. An important issue arises from the fact that the CMS and DA data provide no basis for discriminating between these different models; all four are equally good as far as these data are concerned. Bell's article makes clear that, in fact, there is a whole family of such models. Thus, in considering the combining of DA and CMS results, it is important to keep in mind that there are many different ways of doing this that are equally consistent with the DA and CMS data but that, nevertheless, produce different subnational estimates.

Unless research on improving DSEs succeeds in bringing CMS national results into essential agreement with DA, then a decision on combining should consider the range of variation across reasonable combining variants. This should be considered in relation to the variation across possible basic CMS estimates, as well as sampling and other uncertainty in such estimates. (This is elaborated below, and further research on both these topics is proposed as activity 6 in Section 6.) In the end, a decision to combine will eliminate the sort of discrepancies between CMS and DA results shown in Figure 1, but must accept the fact that combining methods other than the one chosen would be equally consistent with the data but would produce different subnational results. We must then prepare to defend the choice of combining method, though it is unlikely we will have data to support this choice over any other. On the other hand, a decision not to combine implicitly accepts the sort of discrepancies shown in Figure 1 between the CMS and DA results at the national level. We would thus need to defend this decision in the presence of DA data suggesting significant CMS undercoverage of adult black males.

There is a range of uncertainty about the basic CMS estimates, arising both from sampling error and from some alternative decisions that could be made about the nature of the CMS estimators (e.g., alternative choices of poststratification). In regard to the latter, the CMS and DA data potentially provide evidence for discriminating between alternative models, either through formal statistical tests, or through more informal, though still data-based, analyses. (See, for example, the discussion in Hogan (1993, pp. 1052-1054) about advantages to the 1990 PES estimates based on 357 poststrata over the original estimates based on 1,392 poststrata.) This is not to say that available data will necessarily discriminate *effectively* between any two alternative CMS estimates; evidence in favor of one or the other approach could be quite weak. We simply note the distinction from the situation with alternative combined DA-CMS estimates, for which the DA-CMS data provide *no* information in favor of one approach over another.

An analogy to regression may help in understanding this issue. Suppose two researchers propose different regression models for a given data set of 100 points. If one model proposed includes 5 variables, and the other includes these 5 variables plus an additional 5 variables, or just a different set of 5 variables, then standard statistical techniques (F-test, nonnested model comparison statistics) can be used to discriminate between them. For two models that involve a large number of variables, say 50 variables each, typically the data will not very effectively discriminate between them. But if the two models proposed involve different sets of 100 variables each, then the data provide no basis for discriminating between the models. Such models are called "saturated models." They have as many parameters as there are data points, which leaves no degrees of freedom for model assessment. This is essentially the situation with combined estimates using the CMS and DA data.

Data-based discrimination of alternative combining variants requires an additional data source beyond the CMS and DA data. The ultimate in this is development of a third data "system" that can be matched to the census and CMS to permit so-called "triple system estimation." The two potential data sources that have been suggested for this purpose are (1) doing two coverage surveys instead of one (e.g., pre-enumeration and post-enumeration surveys), or (2) using administrative records data (studied in the 1988 test census). Research on these approaches is listed among the projects considered in Section 7 (see activity 1.d there), but we give it low priority because we feel both approaches have significant feasibility problems. Alternatively, if some subnational population estimates or indicators can be developed from admininstrative records (short of a matchable third system), this may help discriminate among combining variants. This requires, however, that such indicators not be subject to the same sorts of biases as the census and CMS results (or to worse biases). Great care must be taken to assure this. It implies, for example, that other information from the decennial census would not be useful for this purpose.

To illustrate the magnitude of variation in subnational estimates over alternative ways of combining 1990 PES (357 poststrata) and DA results, Figure A-1 compares 50 state results from the four models considered in Bell (1993). The results displayed are state census undercount rates (100(PES_c - Census)/PES_c) for total pop) obtained from the four models, where PES_c is any of the four combined PES-DA estimates. These are contrasted with the undercount rates obtained from the basic PES 357 poststrata results (usual DSEs without regression smoothing). The four combined alternatives produce higher undercount rates (and therefore, larger adjustments to the census counts), for all states. The range of variation among the four alternative DSEs seems small for some states, but not so small for others. Relations between the alternatives vary across states because they all control adult male estimates to the same national totals, but do so in different ways.

Apportionment results for the House of Representatives were also obtained for the 357 poststrata DSEs and the combined estimates. Relative to the unadjusted census, the PES with 357 poststrata makes one change: California gains a seat and Wisconsin loses a seat. The four combined estimators all make one additional change: New York gains a seat that is lost by Pennsylvania.

For comparison, Figure A-2 shows state undercount rates from two alternative DSEs - those from 357 poststrata as shown in Figure A-1, and a set of revised DSEs from 1,392 poststrata (taken from the April 24, 1992 memo from John Thompson to the CAPE committee and working group that presented the 357 poststrata results.) Figure A-2 shows generally as much or more difference between these two results as there is variation among the alternative combined estimates in Figure A-1. The comparison in Figure A-2 needs to be qualified, however. The 1,392 poststrata results include the edit corrections that affected original estimates of erroneous enumerations, but do not include the results of reworking the matching results for the 104 most influential block clusters (which were used for the 357 poststrata results). There is thus an unknown degree of noncomparability to these results; it would be preferable to have results for the 1,392 poststrata with the reworked matching results for comparison. Nevertheless, this at least provides some information by showing the range of variation in alternative versions of basic DSEs that were considered. Also relevant are the standard errors of the 357 poststrata state undercount rates (reported in the April 24, 1992 memo.) These generally range between .3 and .5 (larger for a few small states), reflecting a range of uncertainty due to sampling error that is of comparable magnitude to the range of variation in the alternative combined estimates.

Figures A-3 and A-4 show analogous comparisons of undercount rates to those of Figures A-1 and A-2, but for 30 large cities. Note first that the range of the vertical scale is about twice that of Figures A-1 and A-2: cities have considerably higher undercount rates than do states. Also, the ranges of variation among the alternative estimates are larger for cities than for states. Standard errors of the 357 poststrata DSEs (from the April 24, 1992 memo) range between .5 and 1.0, reflecting more sampling error uncertainty. Thus, relative comparisons of Figures A-3 and A-4 lead to similar conclusions as those from Figures A-1 and A-2. Again, it would be desirable (for Figure A-4) to instead show results for 1,392 poststrata with the reworked matching results.

It should be noted that no research on combining methodology, or on improving the DA estimates, can resolve the issue that different ways of combining CMS and DA results will produce different subnational estimates. Neither is this issue addressed by most of the research projects on statistical estimation discussed in Section 7. Only the projects listed there under general activity 1, which involve investigating ways to reduce correlation bias in DSEs, and thus bring CMS estimates closer to DA, attempt to address this problem. To the extent that such research is successful, it could reduce the importance of this issue. Activity 6 in Section 6, which involves studying the range of variation across reasonable combining variants, attempts to assess the magnitude of this problem.

Essentially the same issue arises in regard to use of DA (or census) population controls for surveys. This is generally accomplished by some form of raking, but different approaches to raking that are equally consistent with the data used would produce different subnational results. In fact, two of the approaches of Bell (1993) and some other proposed combining approaches can be viewed as forms of raking, since they allocate the discrepancy between DA and CMS results at the national level (by age-race groups) among subnational poststrata proportional to some associated size measure (e.g., estimated census omissions). The main difference between survey raking to population controls, versus combining CMS and DA results in census estimation, may be simply that the stakes are much higher in the census.

Figure A.1 1990 State Undercount Rates from Alternative DSEs

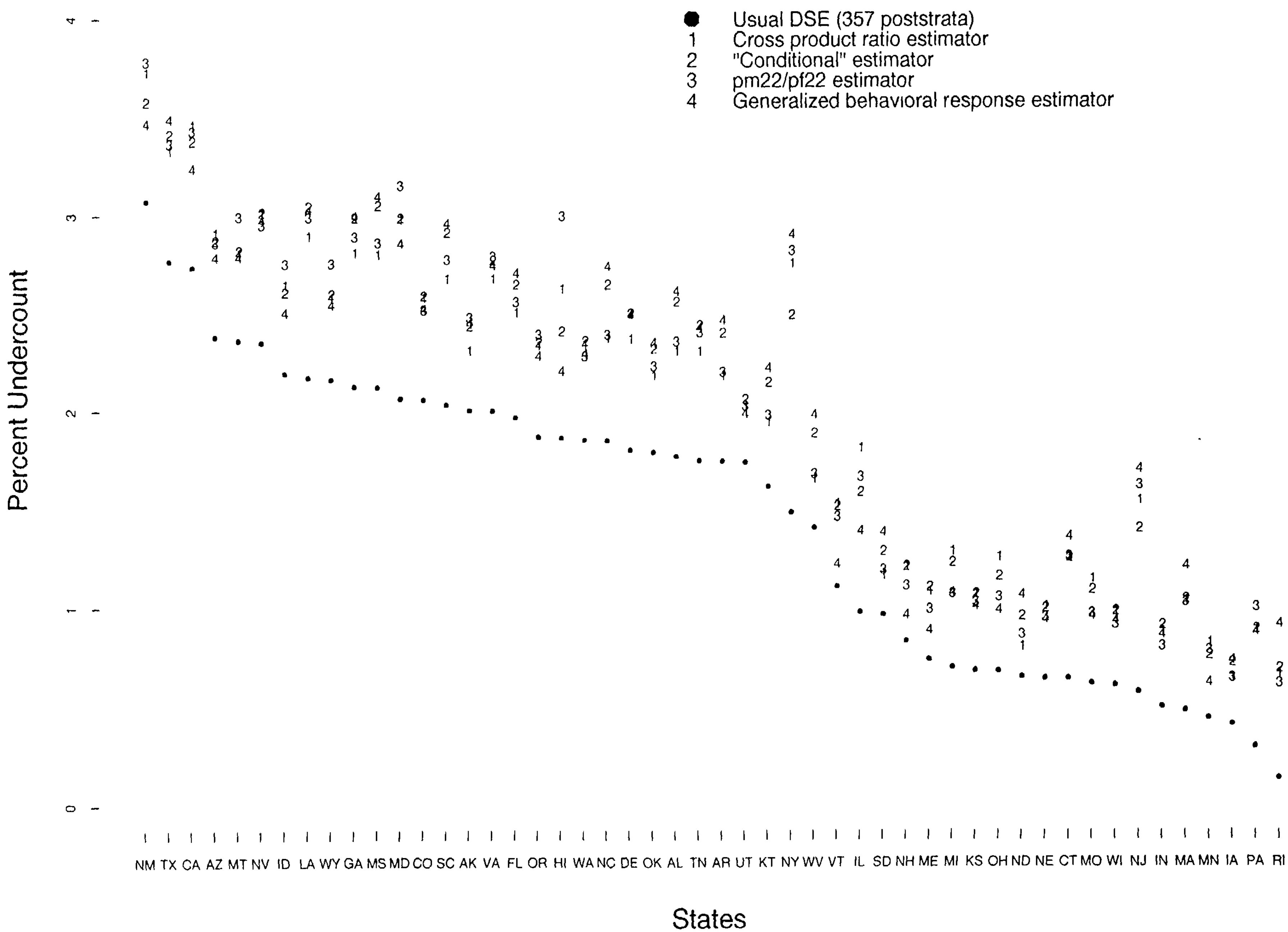


Figure A.2 1990 State Undercount Rates from Alternative DSEs

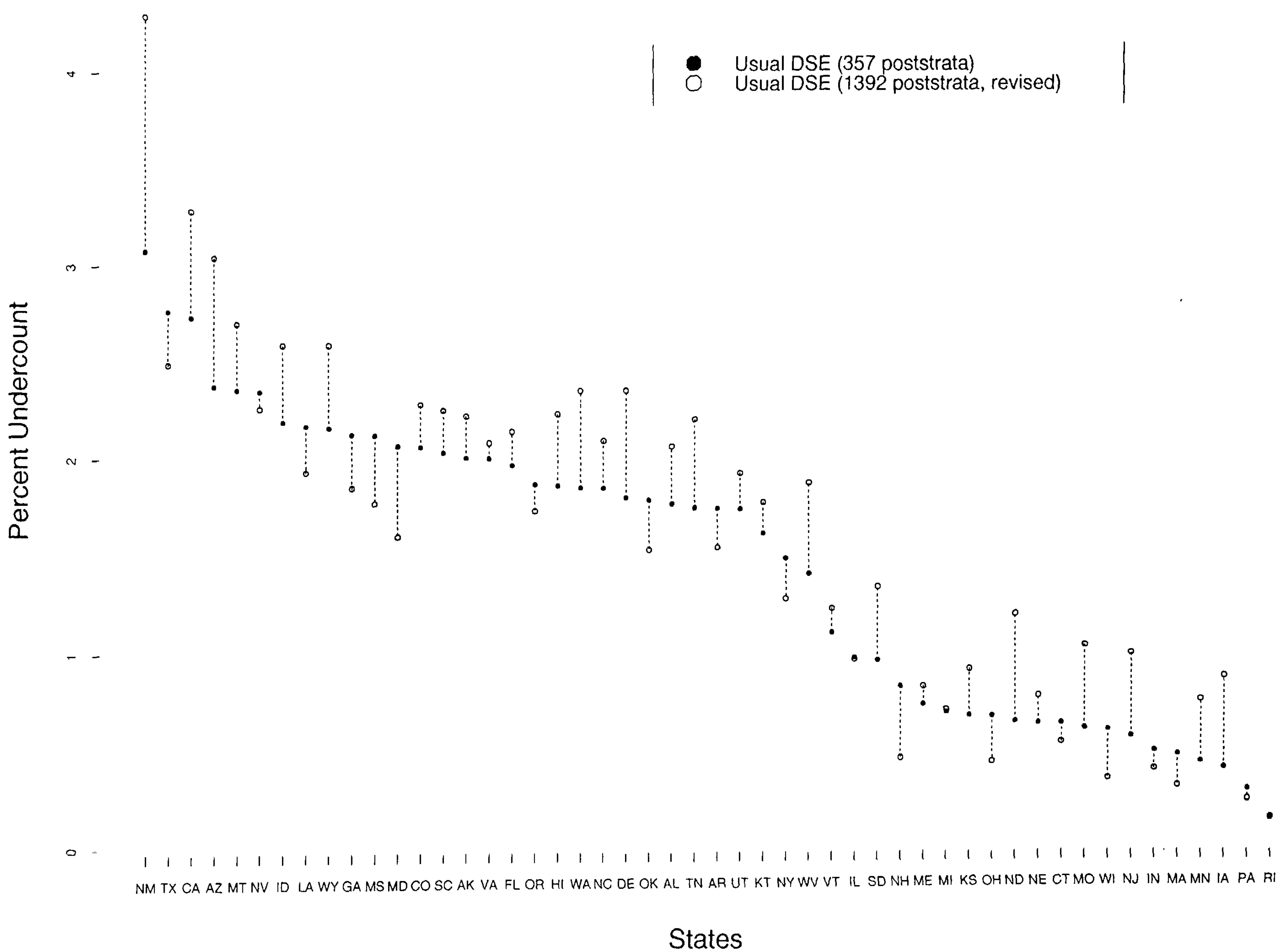


Figure A.3 1990 City Undercount Rates from Alternative DSEs

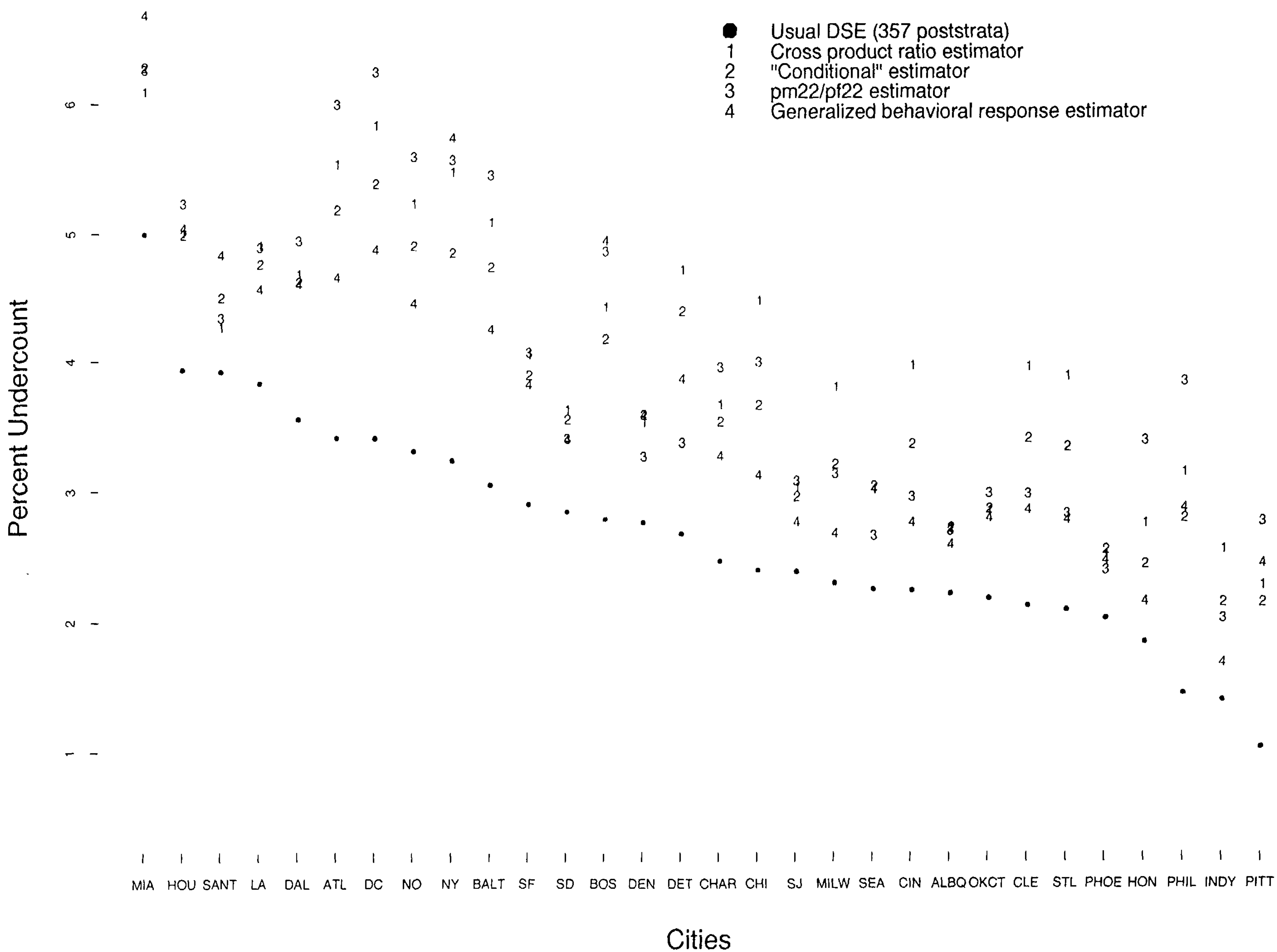
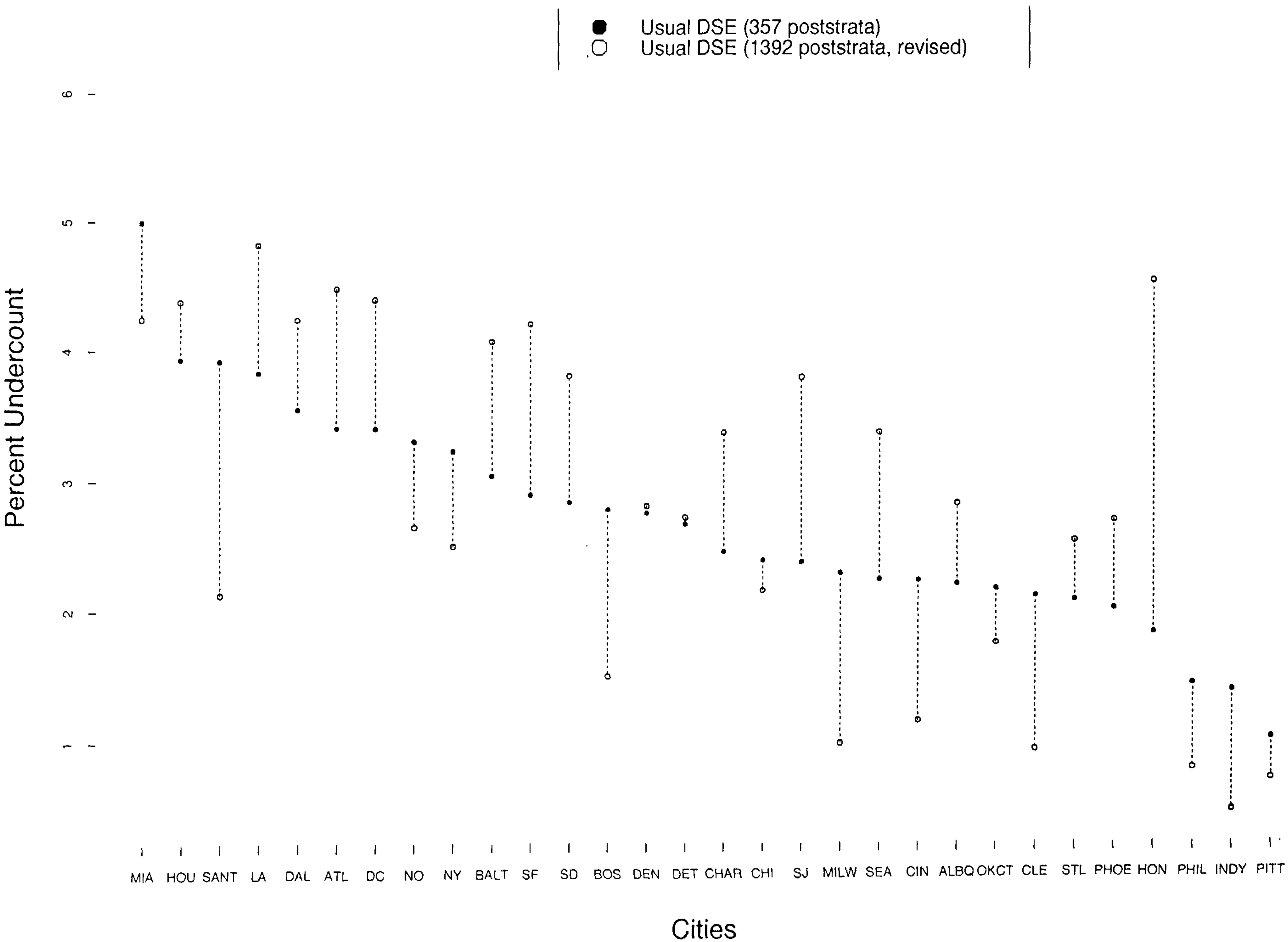


Figure A.4 1990 City Undercount Rates from Alternative DSEs



APPENDIX B

Demographically Meaningful Ways to Combine Yield Similar Subnational Results

With regard to the four different models that Bell (1993) developed, he states that "... the CMS and DA data provide no basis for discriminating between these different models; all four are equally good as far as these data are concerned. ...Thus, in considering the combining of DA and CMS results, it is important to keep in mind that there are many different ways of doing this that are equally consistent with the DA and CMS data but that, nevertheless, produce different subnational estimates."

It is true that we do not have independent observed values for the post-stratum males and, therefore, cannot compare them with the fitted (model-based) values by using a statistical test such as a χ^2-test of goodness of fit. However, it does not necessarily follow that there is no meaningful way to compare different models.

A standard component of judging models is to use a set of postulates (i.e., reasonable assumptions) and then study how different models fare in terms of these postulates. The two postulates implied in Bell's models are:

1) The dual-system estimates (DSE's) of females for the post-strata within an age-race group based on the assumption of census-CMS independence are correct,

2) The DA sex-ratios by age and race (at the national level) are correct.

With these two postulates, Bell encountered numerous possible ways of combining and came to the conclusion that with combined CMS-DA estimates, "We must... prepare to defend the choice of combining method, though it is unlikely we will have data to support this choice over any other."

Additional Postulates

If there are other postulates to add to the two given above, it may be possible to eliminate some of the combining methodologies that Bell considered. It may even be possible to get to the point where we are left with only a very small number (hopefully one) of meaningful models that would satisfy all the postulates.

In order to make the results from the models demographically meaningful, the following three postulates should be considered for addition to the two listed above. This presentation does not claim that the five postulates are in order of defensibility or that the list is exhaustive.

3) Because of the appearance of negative numbers in the individual cells in the 2x2 tables of the census-CMS matched data, models involving individual cell numbers with negative values are expected to give implausible results, and, therefore, should be avoided.

Ideally and conceptually, the 2x2 cells cannot have negative numbers. However, as we see in Table B-1, some (1,2)-cell numbers are negative, and because of that, the corresponding (2,2) cells and the (.,2) cells based on the dual-system model are also negative. Until this problem with the CMS data is resolved, we cannot ignore this point when we propose a model. Otherwise, we may have a demographically unrealistic situation of a modified census undercount rate[1] of -33.3 percent for males (Table B-3, Bell's Model 3). Replacement of the negative numbers by zeros is often suggested, but that does not satisfactorily solve the problem.

4) The estimated numbers for males for the post-strata within an age-race group is not totally independent of the corresponding estimated numbers for females, and, therefore, the models should involve the female DSE's.

This is similar to the concept of using auxiliary correlated variables in the ratio and regression estimates. Without this male-female linkage, we may have a demographically implausible situation where the male and female modified undercount rates in a post-stratum are, respectively, 1.6 and 15.9 percent, and the corresponding rates in another post-stratum are 40.9 and 11.9 percent (Table B-3, Bell's Model 1). For the same reason, we may also have an adjusted sex-ratio of .821 which is even less than in the census (Table B-3, Bell's Model 4).

5) Both the CMS and the census data for males should be used for the male models, i.e., we should not ignore either of the two sources of data.

Bell states that the combined estimates using the CMS and DA data may be regarded as coming from "saturated models" having as many parameters as there are data points and leaving no degrees of freedom for model assessment. This is true if we have only the male 2x2 tables for the post-strata with missing fourth cells along with the known sum of these missing fourth cells. However, this scenario changes drastically as soon as we assume complete knowledge of female numbers (Postulate 1 by Bell) but fail to use it either in the models (Bell's Models 1, 2, and 4) or in the demographic interpretation of the results from the models. Postulate 4 attempts to rectify this problem.

[1]Modified census undercount rate is defined here as census undercount rate based on the census number and the estimated number, both taken from the census-CMS matched 2x2 table. A modified rate is used in this analysis because data on erroneous enumerations and imputed numbers in the census needed to calculate the census undercount rate, as usually defined, are excluded from the 2x2 table. The conclusions, however, remain unchanged irrespective of how the undercount rate is computed.

As noted earlier, the five postulates above are not necessarily arranged in order of defensibility. For example, some analysts may regard the assumption of no error (e.g., no correlation bias) for females in Postulate 1 as less defensible than the assumption of male-female linkage in Postulate 4.

Bell's Models

Bell's four models can be reviewed with the three additional postulates (Nos. 3 to 5).

Model 1: In terms of the symbols in Table 1, this model uses the equation

$$\frac{M_{11}\hat{M}_{22}}{M_{12}M_{21}} = C \qquad \ldots(1)$$

for all post-strata, where the constant C is determined from the assumption that, at the national level, the total DSE for females multiplied by the DA sex-ratio gives the total number of males, i.e.,

$$\sum \hat{M}_{22} = \left(\sum F_{..}\right) x\ DA\ Sex\text{-}Ratio - \sum (M_{11}+M_{12}+M_{21}) \quad . \qquad \ldots(2)$$

Model 2: This model requires that for all post-strata,

$$\frac{M_{11}(M_{21}+\hat{M}_{22})}{M_{1.}M_{21}} = C \quad , \qquad \ldots(3)$$

subject to the constraint in (2).

Model 3: This model estimates the missing cell for all post-strata from the equation

$$\frac{\hat{M}_{22}F_{..}}{(\hat{M}_{22}+M_{11}+M_{12}+M_{21})\,F_{22}} = C \quad , \qquad \ldots(4)$$

where C is determined from equation (2).

Model 4: This model uses the equation

$$\frac{M_{21}(\hat{M}_{22}+M_{11}+M_{12}+M_{21})}{M_{1.}(M_{21}+\hat{M}_{22})} = C \quad , \qquad \ldots(5)$$

subject to the same constraint in (2).

Models 1 and 3 use, respectively, M_{12} and F_{22} (DSE), both of which can be negative (see Table B-1). Therefore, these two models do not satisfy postulate 3.

Models 1, 2, and 4 do not involve the female numbers, i.e., they are developed independently of the female dual-system estimates (based on census-CMS independence). Therefore, these models fail to satisfy postulate 4. Only Model 3 satisfies this postulate.

All four models use the census-CMS matched data. Therefore, all of them satisfy postulate 5.

In sum, none of Bell's four models satisfies all five of the postulates.

Proposed Model

The proposed model is a very simple one which says that the ratio of the modified census-CMS coverage rate[2] for males (based on the proposed model results) to the modified census-CMS coverage rate for females (based on the DSE) is a constant for all post-strata. In symbols,

$$\frac{\dfrac{M_{11}+M_{12}+M_{21}}{M_{11}+M_{12}+M_{21}+\hat{M}_{22}}}{\dfrac{F_{11}+F_{12}+F_{21}}{F_{..}}} = C \ , \qquad \ldots(6)$$

subject to the constraint in (2).

The model in (6) satisfies all five of the postulates. In fact, there may not be any other simple and meaningful model for males which can compete with this model.

Illustration

Tables B-1 to B-4 illustrate the application of Bell's four models and the proposed model to 12 post-strata for Blacks aged 30-49 in 1990. Table B-1 gives the underlying DSE numbers, Table B-2 gives the estimated males from the five models, Table B-3 shows the modified census undercount rates and sex-ratios, and Table B-4 provides the modified census-CMS coverage rates based on these models.

The modified census undercount rates (in percent) for males from Bell's models (Table B-3) show significant variations both across models and across post-strata. For example, for post-stratum 4, the modified undercount rates are 1.6, 18.6, -33.3 and 30.3 from the four models; and, for Model 3, the modified undercount rates in the post-strata range from -33.3 to 50.3. The post-stratum modified undercount rates for males from the proposed model, on the other hand, show more plausible variation (17.9 to 34.3) around the overall rate of 24.1.

[2] As in the case of modified undercount rate, modified coverage rate excludes the erroneous enumerations and the imputed numbers from the total persons in the census.

In spite of the fact that the overall sex-ratio increases from .782 (from the census) to .906 (from the models), the sex-ratios for some of the post-strata from Bell's models are, in fact, lower than those from the census: post-stratum 4 in Model 1 and Model 3, and post-stratum 1 in Model 4. The sex-ratios in the census range from .623 to .929 whereas those from Bell's models show a much wider range (from .586 to 1.242). On the other hand, the proposed model gives higher sex-ratios than those from the census for all post-strata, and shows a more plausible range (.739 to 1.050).

Table B-4 depicts more clearly the results from Bell's models and compares them with the results from the proposed model in terms of the modified coverage rates of combined census and CMS from the matched data in the 2x2 tables. Using the symbols in Table B-1, this rate for males, for example, is defined as

$$\frac{M_{11}+M_{12}+M_{21}}{M_{11}+M_{12}+M_{21}+\hat{M}_{22}} \; x\ 100 \quad ,$$

where $\hat{M}_{22}$ is the estimated number for the missing cell from the model.

The unrealistic patterns of male and female modified coverage rates for the 12 post-strata within Blacks aged 30-49 from Bell's models are evident from the following comparisons in Table B-4:

Bell's Model	Post-Strata	Female	Male	Male/Female Ratio
1	4	110.5	120.3	1.088
	9	99.5	75.3	0.756
2	5	100.2	95.7	0.955
	10	98.5	79.4	0.806
3	4	110.5	162.8	1.473
	7	94.0	63.9	0.681
4	1	96.3	95.8	0.995
	4	110.5	85.1	0.770

The two ratios forming the pairs in the last column of the above table are very different. We expect that within an age-race group, the ratios of the modified coverage rates for males and females in different post-strata would be approximately equal, reflecting Postulate 4. As we see in Table B-4, all 12 ratios corresponding to the proposed model are 0.899.

We should note here that even if male-female linkage (Postulate 4) is a reasonable assumption, the proposed model may produce unacceptable results if the assumption of correct DSE's for females (Postulate 1) is not a reasonable one. For example, the modified coverage rates from the female DSE's in Table B-4 for Blacks, 30-49, in other urban areas are 100.2 for owners (post-stratum 5) and

101.5 for non-owners (post-stratum 11). If this higher coverage rate for non-owners than for owners poses a problem (i.e., if we question Postulate 1), this problem is also transmitted, through Postulate 4, to the corresponding male modified coverage rates of 90.0 and 91.2 in the proposed model.

In conclusion, the proposed model may not be the best possible model for the census-CMS-DA combined estimates for males. However, if we have a number of models all satisfying the five postulates given earlier, we would expect the results from these models to be very close. The results in Figures A-1 and A-3 based on Bell's models show substantial differences. These differences are largely due to the fact that negative numbers were involved in Models 1 and 3 (which would be inconsistent with Postulate 3) and female numbers were not used for Models 1, 2, and 4 (which would be inconsistent with Postulate 4).

Table B-1. Dual-System Estimates (DSE's) for 12 Post-Strata Based on the CMS and the Census, for Blacks Aged 30-49: United States, 1990

Post*-Strata	Census and CMS	Census, Not CMS	Total Census **	CMS, Not Census	Not CMS, Not Census (DSE)	Total Not Census (DSE)	Total CMS	Total Not CMS (DSE)	Total (DSE)
					Male				
	M_{11}	M_{12}	$M_{1.}$	$M_{21.}$	M_{22}	$M_{2.}$	$M_{.1}$	$M_{.2}$	$M_{..}$
1	141567	64115	205682	17672	8003	25675	159239	72118	231357
2	379139	81709	460848	40835	8800	49635	419974	90509	510483
3	210725	49562	260287	23433	5511	28944	234158	55073	289231
4	178356	-53397	124959	27737	-8304	19433	206093	-61701	144392
5	314080	1519	315599	33038	159	33197	347118	1678	348796
6	246492	32307	278799	38469	5042	43511	284961	37349	322310
7	216082	84823	300905	86040	33777	119817	302122	118600	420722
8	279818	149868	429686	84300	45149	129449	364118	195017	559135
9	136885	78691	215576	48690	27992	76682	185575	106683	292258
10	111513	57801	169314	44969	23308	68277	156482	81109	237591
11	313894	-16237	297657	67312	-3481	63831	381206	-19718	361488
12	52785	28097	80882	11778	6269	18047	64563	34366	98929
Total	2581336	558858	3140194	524273	152225	676498	3105609	711083	3816692
					Female				
	F_{11}	F_{12}	$F_{1.}$	F_{21}	F_{22}	$F_{2.}$	$F_{.1}$	$F_{.2}$	$F_{..}$
1	173719	75143	248862	24451	10578	35029	198170	85721	283891
2	502134	44902	547036	35945	3215	39160	538079	48117	586196
3	271782	36410	308192	22838	3060	25898	294620	39470	334090
4	223704	-89147	134557	42301	-16857	25444	266005	-106004	160001
5	371117	-10109	361008	25382	-692	24690	396499	-10801	385698
6	274158	44822	318980	39471	6453	45924	313629	51275	364904
7	328988	121251	450239	95224	35096	130320	424212	156347	580559
8	395789	179870	575659	59477	27031	86508	455266	206901	662167
9	333752	12198	345950	51880	1896	53776	385632	14094	399726
10	177245	25943	203188	23922	3501	27423	201167	29444	230611
11	482488	-66636	415852	50058	-6913	43145	532546	-73549	458997
12	77267	28150	105417	8457	3082	11539	85724	31232	116956
Total	3612143	402797	4014940	479406	69450	548856	4091549	684255	4563796

* Owner: Urbanized Areas 250,000+ (1 = North East, 2 = South, 3 = Midwest, 4 = West)
Other Urban (5), Non-Urban (6).
Non-Owner: Urbanized Areas 250,000+ (7 = North East, 8 = South, 9 = Midwest, 10 = West)
Other Urban (11), Non-Urban (12).
** = (Census count excluding imputation) x [1 - (weighted erroneous enumerations/weighted E-sample total)]

Table B-2. Census, CMS, and DSE Populations by Sex, and Male Populations Estimated by 5 Models Using DA Sex-Ratio[2], for Blacks Aged 30-49: United States, 1990

Post-Strata[1]	Female			Male			Males Estimated by Bell's 4 Models				Males by Proposed Model
	Census	CMS	DSE	Census	CMS	DSE	1	2	3	4	
1	248862	198170	283891	205682	159239	231357	248155	243483	287163	233054	258129
2	547036	538079	586196	460848	419974	510483	528953	533924	518646	524360	561267
3	308192	294620	334090	260287	234158	289231	300798	302900	300112	296814	318594
4	134557	266005	160001	124959	206093	144392	126964	153569	93777	179390	153701
5	361008	396499	385698	315599	347118	348796	349132	364474	344946	368302	387210
6	318980	313629	364904	278799	284961	322310	332892	342858	354671	343608	359358
7	450239	424212	580559	300905	302122	420722	491604	477303	605079	493986	458229
8	575659	455266	662167	429686	364118	559135	653894	620269	679374	587149	596216
9	345950	385632	399726	215576	185575	292258	351000	328469	271959	311860	295432
10	203188	201167	230611	169314	156482	237591	286511	269836	235614	265823	242094
11	415852	532546	458997	297657	381206	361488	354180	391631	334890	430839	400051
12	105417	85724	116956	80882	64563	98929	112087	107452	109936	100983	105887
Total	4014940	4091549	4563796	3140194	3105609	3816692	4136168	4136168	4136168	4136168	4136168

[1] For definition of post-strata, see footnote of Table B-1.
[2] DA sex-ratio for Blacks, 30-49 (excluding military and institutional population) for 1990 = 0.90630.

Table B-3. Modified Census Undercount Rates and Sex-Ratios Based on DSE's and 5 Other Models, for Blacks Aged 30-49: United States, 1990

Post-Strata**	Modified Census Undercount Rates (Percent)*** DSE Female	DSE Male	Male (Bell's Models) 1	2	3	4	Male Proposed Model	Sex-Ratios (males per female) Census	CMS	DSE	Bell's Models 1	2	3	4	Proposed Model
1	12.3	11.1	17.1	15.5	28.4	11.7	20.3	.826	.804	.815	.874	.858	1.012	.821*	.909
2	6.7	9.7	12.9	13.7	11.1	12.1	17.9	.842	.781	.871	.902	.911	.885	.895	.957
3	7.8	10.0	13.5	14.1	13.3	12.3	18.3	.845	.795	.866	.900	.907	.898	.888	.954
4	15.9	13.5	1.6	18.6	-33.3	30.3	18.7	.929	.775	.902	.794*	.960	.586*	1.121	.961
5	6.4	9.5	9.6	13.4	8.5	14.3	18.5	.874	.875	.904	.905	.945	.894	.955	1.004
6	12.6	13.5	16.2	18.7	21.4	18.9	22.4	.874	.909	.883	.912	.940	.972	.942	.985
7	22.4	28.5	38.8	37.0	50.3	39.1	34.3	.668	.712	.725	.847	.822	1.042	.851	.789
8	13.1	23.2	34.3	30.7	36.8	26.8	27.9	.746	.800	.844	.988	.937	1.026	.887	.900
9	13.5	26.2	38.6	34.4	20.7	30.9	27.0	.623	.481	.731	.878	.822	.680	.780	.739
10	11.9	28.7	40.9	37.3	28.1	36.3	30.1	.833	.778	1.030	1.242	1.170	1.022	1.153	1.050
11	9.4	17.7	16.0	24.0	11.1	30.9	25.6	.716	.716	.788	.772	.853	.730	.939	.872
12	9.9	18.2	27.8	24.7	26.4	19.9	23.6	.767	.753	.846	.958	.919	.940	.863	.905
Total	12.0	17.7	24.1	24.1	24.1	24.1	24.1	.782	.759	.836	.906	.906	.906	.906	.906

** For definition of post-strata, see footnote of Table B-1.

* Sex ratio is lower than in the census.

*** For better comparison of results from different models, the modified census undercount rates use the census numbers from Tables B-1 and B-2. If the total census counts including erroneous enumerations and imputed numbers are used, the undercount rates would be significantly lower. For example, the undercount rates based on DSE would then come down from 17.7 and 12.0 to 6.8 and 3.2 for males and females, respectively.

Table B-4. Modified (Census + CMS) Coverage Rates for Females Based on DSE's and for Males Based on DSE's and 5 other Models, for Blacks Aged 30-49: United States, 1990

Post-Strata ***	Modified (Census + CMS) Coverage Rates (Percent)							Ratio of Male Modified Coverage Rate to Female Modified Coverage Rate from DSE					
	DSE		Male (Bell's Models)				Male		Bell's Models				
	Female	Male	1	2	3	4	Proposed Model	DSE	1	2	3	4	Proposed Model
1	96.3	96.5	90.0	91.7	77.8	95.8	86.5	1.003**	.935	.953	.808	.995**	.899
2	99.5	98.3	94.8	94.0	96.7	95.7	89.4	.988	.954	.945	.973	.962	.899
3	99.1	98.1	94.3	93.7	94.5	95.6	89.1	.990	.952	.945	.954	.965	.899
4	110.5	105.8	120.3	99.4	162.8	85.1	99.3	.957	1.088**	.900	1.473**	.770*	.899
5	100.2	100.0	99.9	95.7	101.1	94.7	90.0	.998	.997	.955**	1.009	.945	.899
6	98.2	98.4	95.3	92.5	89.5	92.3	88.3	1.002	.970	.942	.911	.940	.899
7	94.0	92.0	78.7	81.1	63.9	78.3	84.4	.979	.838	.863	.681*	.834	.899
8	95.9	91.9	78.6	82.9	75.7	87.5	86.2	.958	.819	.864	.789	.913	.899
9	99.5	90.4	75.3	80.5	97.2	84.7	89.5	.909*	.756*	.808	.976	.851	.899
10	98.5	90.2	74.8	79.4	90.9	80.6	88.5	.916	.759	.806*	.923	.819	.899
11	101.5	101.0	103.0	93.2	109.0	84.7	91.2	.995	1.015	.918	1.074	.835	.899
12	97.4	93.7	82.7	86.2	84.3	91.8	87.5	.962	.849	.886	.866	.942	.899
Total	98.5	96.0	88.6	88.6	88.6	88.6	88.6	.975	.899	.899	.899	.899	.899

* Lowest in the column.
** Highest in the column.
*** For definition of post-strata, see footnote of Table B-1.

APPENDIX C

Specific Research Projects for Improving Demographic Analysis Estimates

Several research activities can be conducted to improve the demographic analysis estimates of population and census coverage. The projects are classified below according to the particular demographic component estimate that would be improved by the research.

Several of the research activities have been carried out before at the Census Bureau (or outside), but the research needs to be continued to maintain the quality of the DA estimates (for example, current estimates of emigration and undocumented immigration). Other research projects focus on validating and improving the estimates (e.g., validation of assumptions about birth registration completeness; research on race classification).

The specific research areas and projects include:

1. Data on births--Births are based on existing administrative records and require essentially no estimation to update the existing historical series to 2000. However, there are several research projects that could be carried out to improve the historical (1935-1990) birth series:

   a) Provide better validation of the assumptions about birth underregistration. This includes assessing the accuracy of the relatively large adjustment factors for the 1930's and 1940's (Census Bureau lead).

   b) Document whether a new test of birth registration completeness is needed (Census Bureau lead).

   c) Investigate the possibility of overregistration of births in counties along the Mexican border (Census Bureau lead)

2. Data on deaths--Deaths are also based on existing administrative records and require no estimation to update the existing historical series to 2000. However, a few research projects can be carried out to improve the historical (1935-1990) death series.

   a) Improve the adjustments for death underregistration for infants (Census Bureau lead).

   b) Investigate the need for an adjustment of underregistration of young children in the 1930's and 1940's (Census Bureau lead).

Demographic Evaluation Projects D1 and D7 provide a comprehensive assessment of the methodology, assumptions, and data sources that underlie the birth and death components as used in the 1990 DA estimates.

3. Data on legal immigration--These estimates are based on virtually complete administrative data sets (from annual INS files). Research to improve this component could include:

a) Improving the period of entry distribution of the historical immigration series. (Census Bureau lead.)

b) Improving the estimates of legal international migration not covered by INS, including Puerto Rican migrants and net movement of civilians overseas (Census Bureau lead).

Demographic Evaluation Project D7 provides a comprehensive assessment of the methodology, assumptions, and data sources that underlie the legal immigration component as used in the 1990 DA estimates.

4. Estimates of undocumented immigration--Research has improved the estimates in the last 15 years, but more needs to be done. Methods for updating and improving the undocumented estimates include those listed below. (The status of past research and the agency or organization that led the research are in parentheses.)

a) Residual estimates. (Methods already developed based on use of 1980 and 1990 census data and CPS data--need refinement and updating with new CPS nativity data: Census Bureau lead.)

b) Consistency correction techniques. (Already developed in analysis of 1950-1980 census data--need to apply to 1990 census data; Urban Institute lead.)

c) Analysis of Visa overstayer data (already developed in analysis of 1986-1992 INS files--needs to be updated; INS lead).

Demographic Evaluation Project D2 provides a comprehensive assessment of the methodology, assumptions, and data sources that underlie the undocumented immigration component as used in the 1990 DA estimates.

5. Estimates of emigration. Methods for improving the emigration estimates include:

a) Residual techniques (method already used in estimating 1960-70 and 1980-90 foreign-born emigration with census data; we are applying the technique to estimate native emigration; Census Bureau lead).

b) Multiplicity estimates (already developed to produce emigration estimates from CPS supplements of 1987, 1988, and 1989--approach needs more research; Census Bureau lead).

c) Use of administrative records (methods being developed; Social Security or other agency lead).

Demographic Evaluation Project D5 provides a comprehensive assessment of the methodology, assumptions, and data sources that underlie the emigration component as used in the 1990 DA estimates.

6. Medicare data--We need to refine the assumptions about underenrollment in the Medicare files, which can be accomplished with research on:

a) Use of cohort component techniques (method already used in 1990 Medicare evaluation--needs to be updated; Census Bureau lead).

b) IRS-Medicare match (done in 1980 and needs to be repeated); Census Bureau lead).

c) Use of survey estimates such as CPS data on Noncash Benefits and SIPP data (methods used in 1990 Medicare evaluation--needs to be updated; Census Bureau lead).

Demographic Evaluation Project D6 provides a comprehensive assessment of the methodology, assumptions, and data sources that underlie the Medicare data component as used in the 1990 DA estimates.

7. Inconsistencies in race classification--The DA estimates will be biased if persons who are classified as Black in DA are reported as another race in the census. We need to conduct research to assess the degree of inconsistency and identify how this "classification error" can be minimized. Specific research projects include:

a) Analyzing 1990 census data on children in racially mixed households. This research may enable us to develop algorithms for assigning race to births of parents of different races (Census Bureau lead).

b) Develop ways to translate demographic race categories into census race categories (in 1990 the census race categories were mapped into DA categories; Census Bureau lead).

c) Conduct a longitudinal match of person records from the 1990 census or current surveys to historical administrative records (e.g., birth certificates, Social Security records, Medicare) to estimate the degree to which race reporting differs between the two sources (study will require extensive research and development and multi-agency participation).

Demographic Evaluation Projects D8 and D9 provide a comprehensive review of the methodology, assumptions, and data sources used to assess the race reporting consistency of the 1990 census and DA estimates.

APPENDIX D

Prioritizing Research on Demographic Analysis: Criteria for Four Different Categories of Impact

(H = high priority, M = medium priority, L = low priority)

1. **Impact on DA estimates of population totals**

H - Research needed because component will comprise large share (> 20 percent) of DA 2000 estimate (see Table 1).

ex. - Births = 84 percent of DA estimate of approx. 277 million

M - Component will comprise moderate share (5 to 20 percent) of DA 2000 estimate, so research needed but has less impact than births.

ex. - Medicare = 13 percent
legal immigration = 8 percent
deaths = 5 percent

L - Component will comprise small share (< 5 percent) of DA 2000 estimate, so research has little impact on overall estimate.

ex. - legal emigration = 2 percent
undocumented immigration = 1 - 2 percent
other migration = < 1 percent)

2,3. **Impact on DA estimates of age-sex structure (2) or race structure (3)**

H - Research on component has potential to significantly alter/refine age-sex or race distributions of the current DA estimates.

ex. - undocumented immigration (age-sex)

M - Research on component has potential to make moderate refinements in the age-sex or race distributions of the current DA estimates.

ex. - legal emigration (age-sex and race), undocumented immigration (race), births (race)

L - Research on component will unlikely lead to any significant modifications of the age-sex or race structure of the current DA estimates.

ex. - births (age-sex)
deaths (age-sex and race)
legal immigration (age-sex and race)
Medicare (age-sex and race)

4. **Impact on measurement of variance**

H - Research has high potential to improve our understanding of the error structure of the component and thus could significantly change current measures of contribution to total variance.

ex. - undocumented immigration
legal emigration

M - Research has some potential to improve our estimates of error in the component, which could lead to changes in the current measures of contribution to variance.

ex. - births
legal immigration
Medicare

L - Research has little potential to significantly change our measurement of error in the component.

ex. - deaths

5. **Research needed because of lack of current estimates**

H - Since little or no current administrative data exists, the estimates for the component need to be updated periodically to account for possible changes in trends (and to make basic improvements in the existing estimates).

ex. - undocumented immigration
legal emigration

M - Current administrative data is available, but some estimation is involved.

ex. - legal immigration (estimation of race distribution, Puerto Rican and civilian citizen migration)

L - Current administrative data is available; little estimation is involved.

ex. - births, deaths, Medicare

APPENDIX E

Uncertainty of DA Proportionate Distributions

Assuming that we will use only the national DA age-sex proportionate distributions for Blacks and Non-Blacks in the integration of DA into the 2000 ICM system, the problem of uncertainties in DA estimates can be posed in an alternative form as follows.

Models with DA age-sex proportionate distributions involve numbers such as the Non-Black females aged 20-29 as a proportion of total Non-Blacks. This proportion was .078 in 1990 DA estimates and we can express this proportion as the sum of products involving various components as shown below:

$$.078 = \frac{\textit{Non-Black females, } 20\text{-}29}{\textit{Total Non-Blacks}}$$

$$= \frac{\sum_{C} \textit{Size of component C in Non-Black females, } 20\text{-}29}{\textit{Total Non-Blacks}}$$

$$= \sum_{C} \frac{\textit{Size of component C in total Non-Blacks}}{\textit{Total Non-Blacks}} \times \frac{\textit{Size of component C in Non-Black females, } 20\text{-}29}{\textit{Size of component C in total Non-Blacks}}$$

=	.754 x .097	(Births)
	- .050 x .042	(Deaths)
	+ .066 x .089	(Legal immigration)
	- .015 x .070	(Legal emigration)
	+ .014 x .133	(Undocumented aliens)
	+ .006 x .143	(Other components)
	+ .130 x .000	(Medicare population)
	+ .095 x .000	(Base population)

In the above 8 product terms representing 8 components, each of the 2nd factors is not expected to have significant error in it. This is because it depends on the age-sex proportionate distribution <u>within</u> each component and is independent of other components. Even if the component is underestimated or overestimated, we do not expect the age-sex proportionate distribution to change significantly.

Therefore, most of the errors in the proportion of Non-Black females, 20-29 (out of total Non-Blacks) lie in the 1st factors, which are proportions of the components in total Non-Blacks. These factors add up to 1, and, therefore, a change in the proportion of one component would change all others.

A simulation or an analytical technique can be developed to assess the error in the proportion of Non-Black females, 20-29 (out of total Non-Blacks) and in all other similar proportions (subject to the constraint that the sum of the first factors is equal to 1). The technique should also take into account another constraint that within each race group, the proportions (such as .078 in the above example) representing the age-sex proportionate distributions should also add up to 1.

APPENDIX F

Subnational DA Estimates for Evaluation of Census 2000

A major focus of the DA program since 1990 has been the development of "subnational" demographic indicators of coverage (including the use of sex ratios). The intent of this new program is to tap into the existing national DA program and population estimates program to provide inexpensive and timely independent estimates for evaluation census and survey-based ICM results. Although this expanded DA program is envisioned for use at the State and large county level, it is already demonstrating its utility in the evaluation of the relative performance of the DSE and CensusPlus survey estimates in the 1995 test sites.

For illustration, demographically-based population and housing unit estimates for subnational areas could serve the following specific purposes:

- in pre-2000 research on stratification and indirect estimation
- assess quality of MAF-based housing unit counts before census begins
- provide early indicators of coverage differentials in the pre-ICM census counts
- allows immediate independent evaluation of survey results (see 1995 ICM Evaluation Projects 14 and 15 for a model of the "subnational" demographic analysis tools that can be used).

The last three items can be thought of as a "continuous count review" program, merging elements of the count review program of the 1990 census with coverage indicators provided by subnational DA. As an example of the first item, research on alternative stratification techniques and indirect estimation models (with 1990 PES data) could use new DA state and county coverage indicators as independent "face validity" benchmarks.

Since the subnational DA program is intended for evaluation of census results--not for integration with ICM--we have not addressed its needed research activities in this report.

Materials attached to these minutes were draft and preliminary material to inform the ESCAP Committee. The data and analysis contained in these documents are subject to revision and are not final. These materials report the results of research and analysis undertaken by Census Bureau staff. They have undergone a more limited review than official Census Bureau publications. Research results and conclusions expressed are those of the authors and do not necessarily indicate concurrence by the Census Bureau.

Revised 2-15-01

Table 3--Estimates of Percent Net Undercount by Sex and Age: 1960 to 2000
(a minus sign denotes a net overcount)

Category	Demographic Analysis					Survey-based	
	1960	1970	1980	1990	2000	PES 1990	A.C.E. 2000
MALE							
Total	3.5	3.4	2.2	2.8	-0.1	1.9	1.5
0-17	2.8	2.7	0.9	2.2	-0.5	3.2	1.5
18-29	5.9	3.9	3.3	2.2	-2.6	3.2	3.5
30-49	4.2	5.1	3.6	3.8	1.3	1.9	1.8
50+	2.2	2.5	1.2	2.7	0.2	-0.6	-0.2
FEMALE							
Total	2.7	2.0	0.3	0.9	-1.2	1.3	0.8
0-17	1.8	2.4	0.9	2.4	0.1	3.2	1.5
18-29	2.8	1.3	0.4	0.6	-3.1	2.8	2.1
30-49	1.9	1.3	-0.0	0.5	-0.9	0.9	1.0
50+	4.6	2.6	-0.2	0.2	-1.4	-1.2	-0.8

Note: DA estimates are consistent with estimates in Table 2.

Revised 2-15-01

Table 4--Estimates of Percent Net Undercount by Race and Sex: 1940 to 2000
(a minus sign denotes a net overcount)

Category	Demographic Analysis						Survey-based	
	1960	1970	1980	1990	2000 Model 1	2000 Model 2	PES 1990	A.C.E. 2000
Total Population	3.1	2.7	1.2	1.8	-0.7	-0.7	1.6	1.15
Black	6.6	6.5	4.5	5.7	4.7	0.9	4.4	2.1
Male	8.8	9.1	7.5	8.5	6.9	3.2	4.9	2.4
Female	4.4	4.0	1.7	3.0	2.5	-1.3	4.0	1.8
Nonblack	2.7	2.2	0.8	1.3	-1.5	-0.9	1.2	1.0
Male	2.9	2.7	1.5	2.0	-1.2	-0.7	1.5	1.4
Female	2.4	1.7	0.1	0.6	-1.7	-1.1	0.9	0.6
Black:Nonblack Diff	3.9	4.3	3.7	4.4	6.2	1.8	3.3	1.0

Note: Model 1 census tabulations for Blacks include persons who marked the Black circle and no other race response Model 2 census tabulations for Blacks include persons who marked the Black circle and other response circles. Persons who marked only the "Other race" circle are reassigned to a specific race category (to be consistent with 1990 DA estimates and the historical demographic data series)

Source: 1940-1990-- Robinson, J. Gregory, Bashir Ahmed, Prithwis Das Gupta, and Karen Woodrow, "Estimates of Population Coverage in the 1990 United States Census Based on Demographic Analysis", Journal of the American Statistical Association, Vol. 88, No. 423, pp. 1061-1077. Estimates for 2000 are unpublished preliminary results.

Source: 2000 - See Table 2. Note that the A.C.E. estimates for Blacks pertain to the Non-Hispanic Blacks

Materials attached to these minutes were draft and preliminary material to inform the ESCAP Committee. The data and analysis contained in these documents are subject to revision and are not final. These materials report the results of research and analysis undertaken by Census Bureau staff. They have undergone a more limited review than official Census Bureau publications. Research results and conclusions expressed are those of the authors and do not necessarily indicate concurrence by the Census Bureau.

Alternate Scenarios for Demographic Components

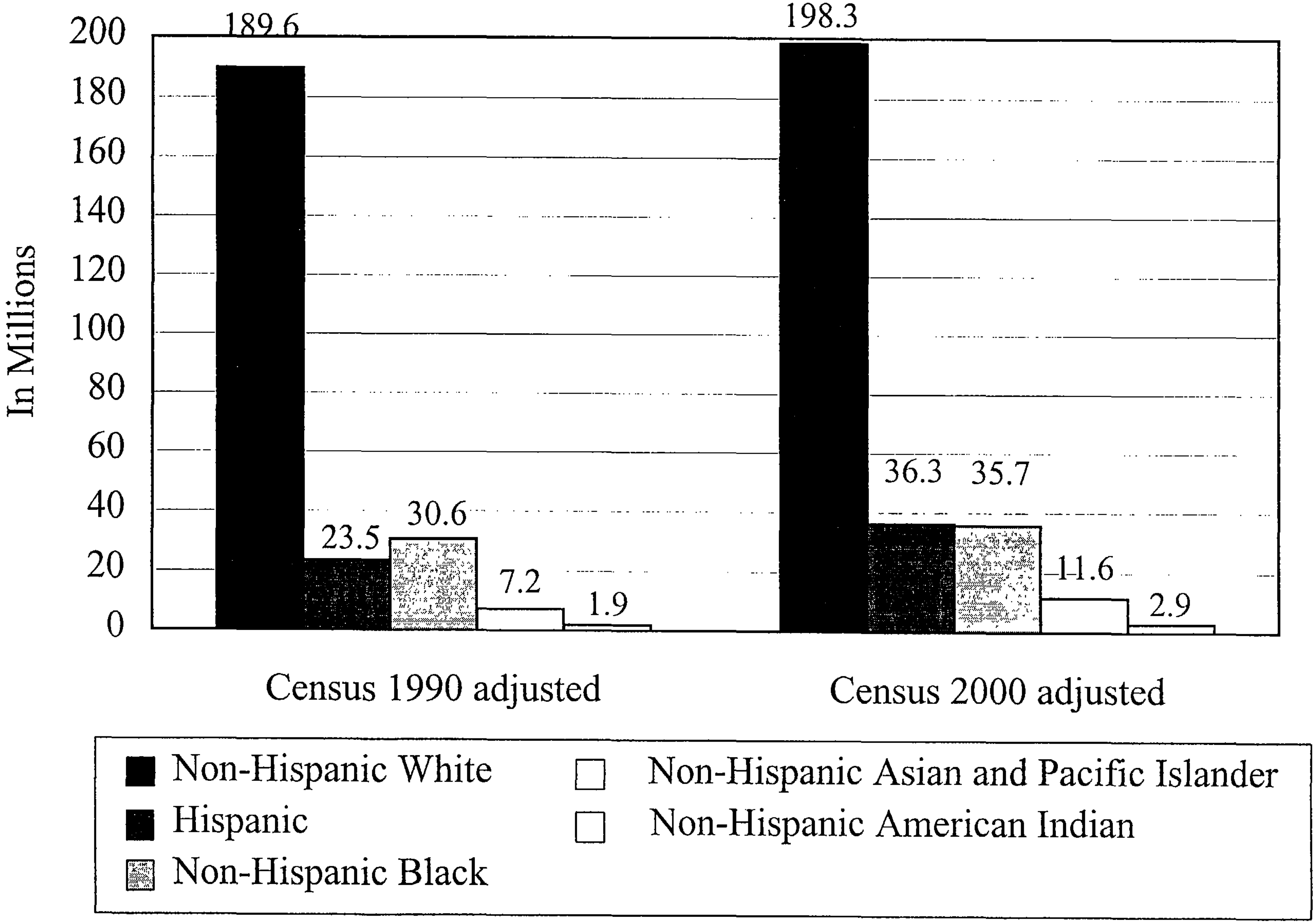
Prepared for the Executive Steering Committee on A.C.E. Policy

February 19, 2001

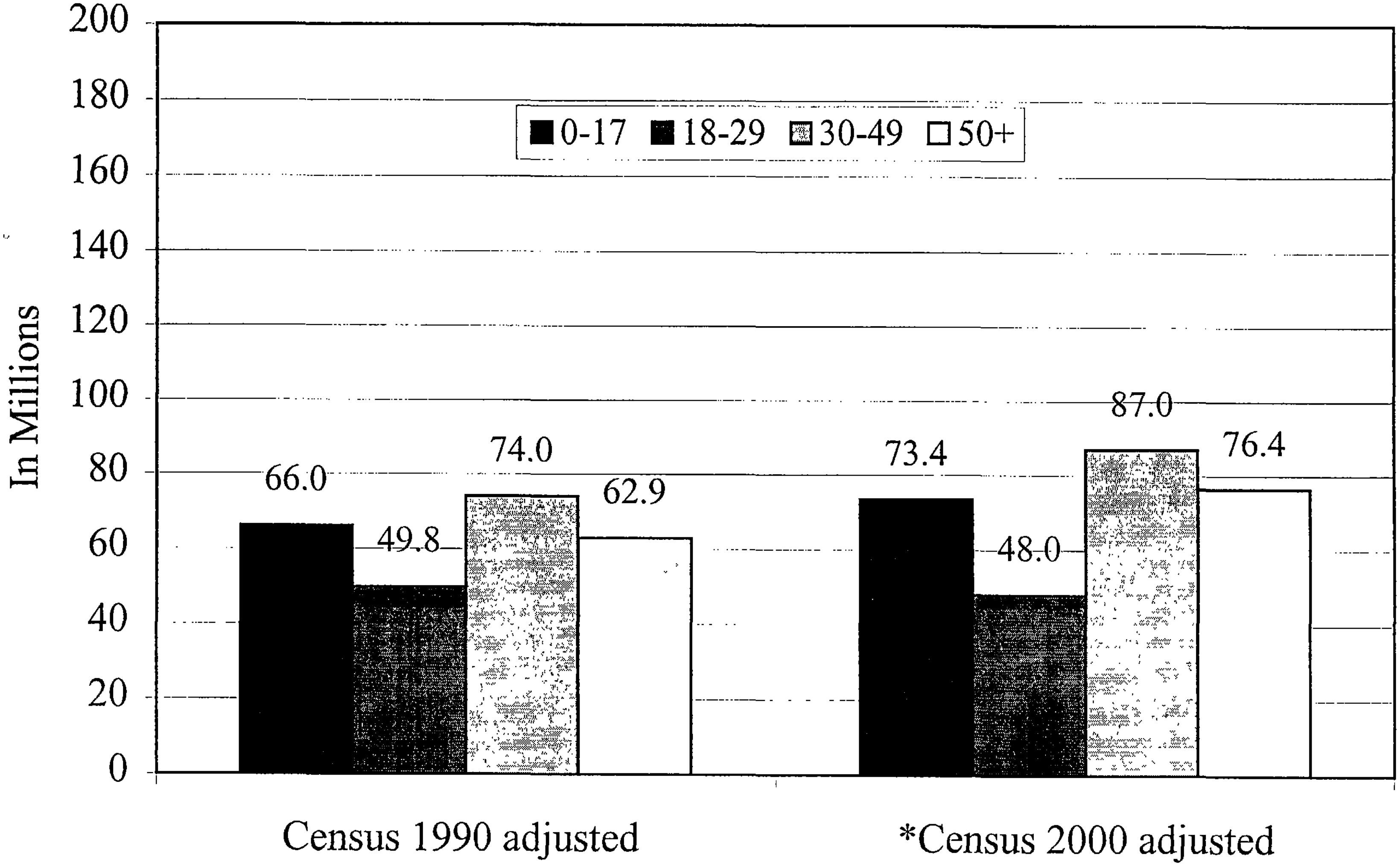
Introduction

- Demographic explanation of change
- Alternate demographic scenarios controlled to Census 2000 adjusted
- Alternate demographic scenarios uncontrolled

Census 1990 (Adj.) and Census 2000 (Adj.) by Race and Hispanic Origin



Census 1990 (Adj.) and Census 2000 (Adj.) by Age



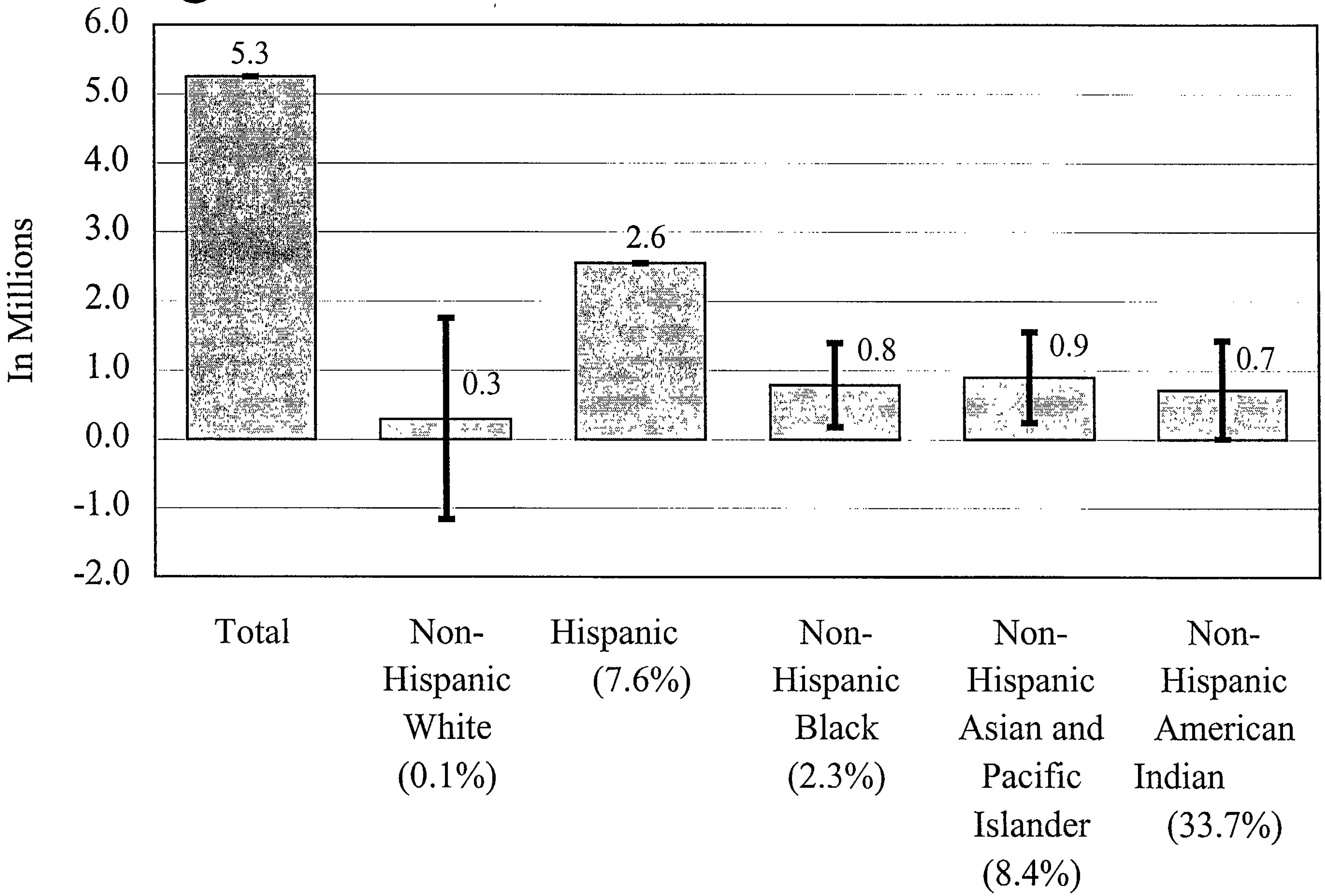
* Calculated using an estimate of the A.C.E.

Population Accounting

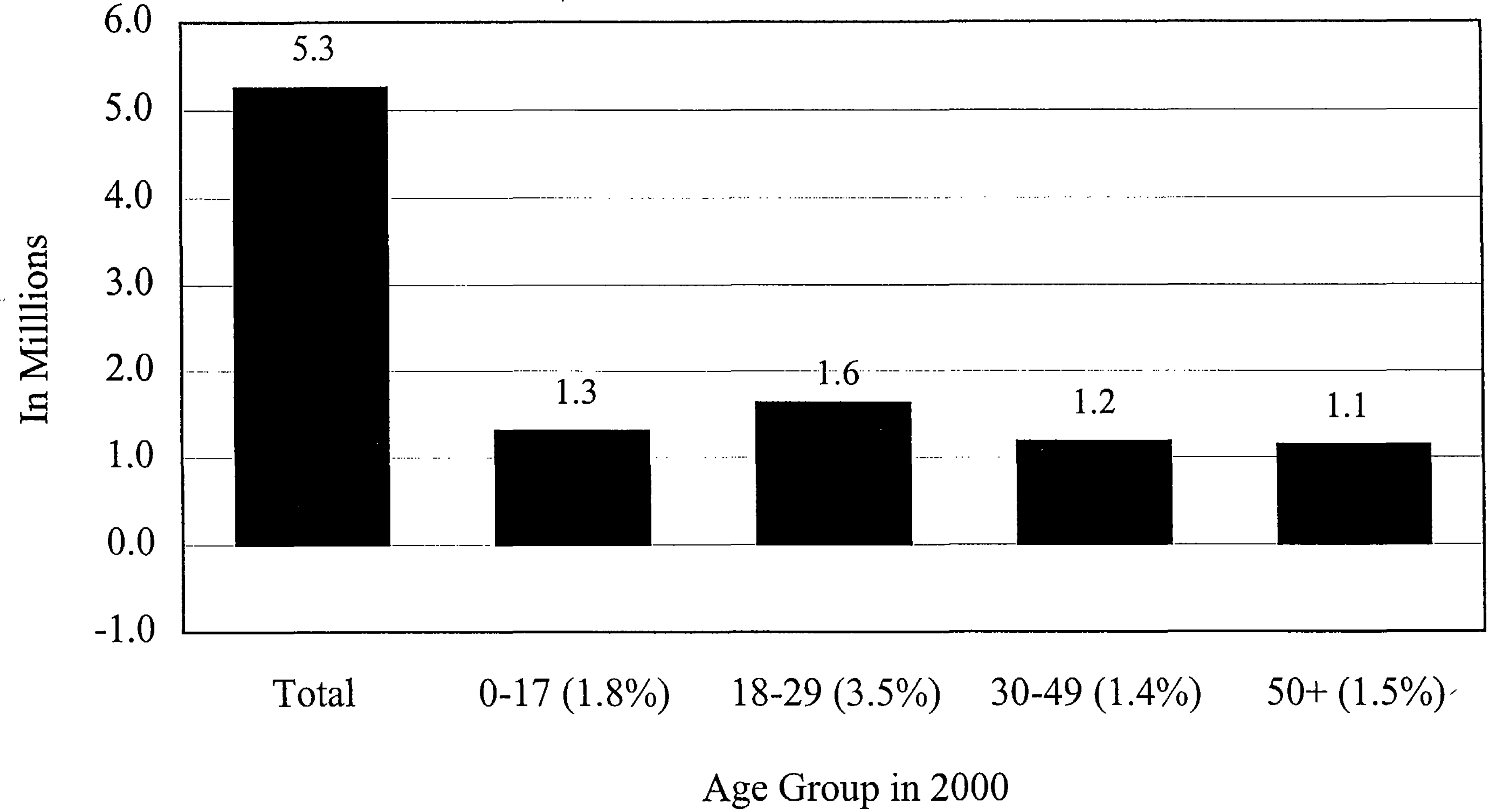
P1990 + PES + B - D + Migration + Unexplained = P2000+A.C.E.

Census 1990 (P1990)		248,700,000
PES (PES)		+ 4,000,000
Births (B)		+ 40,100,000
Deaths (D)		- 22,800,000
Net International Migration (Migration)		+ 9,400,000
Unexplained (Unexplained)		+ **5,300,000**
Census 2000 (P2000)	281,400,000	
A.C.E.	+ 3,300,000	
Adjusted Census 2000		284,7000,000

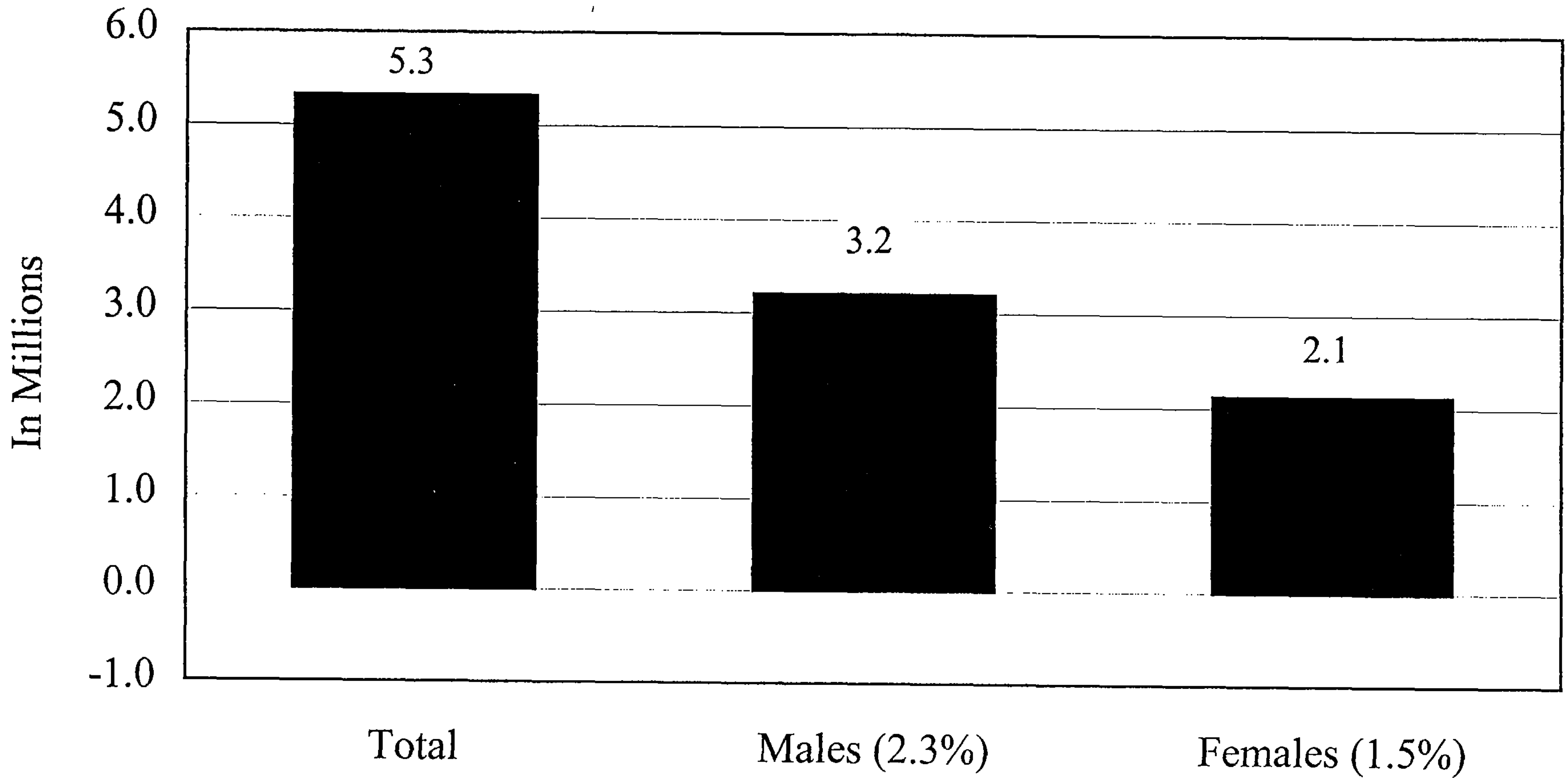
Unexplained Change by Race and Hispanic Origin with Minimums and Maximums



Unexplained Change by Age



Unexplained Change by Sex

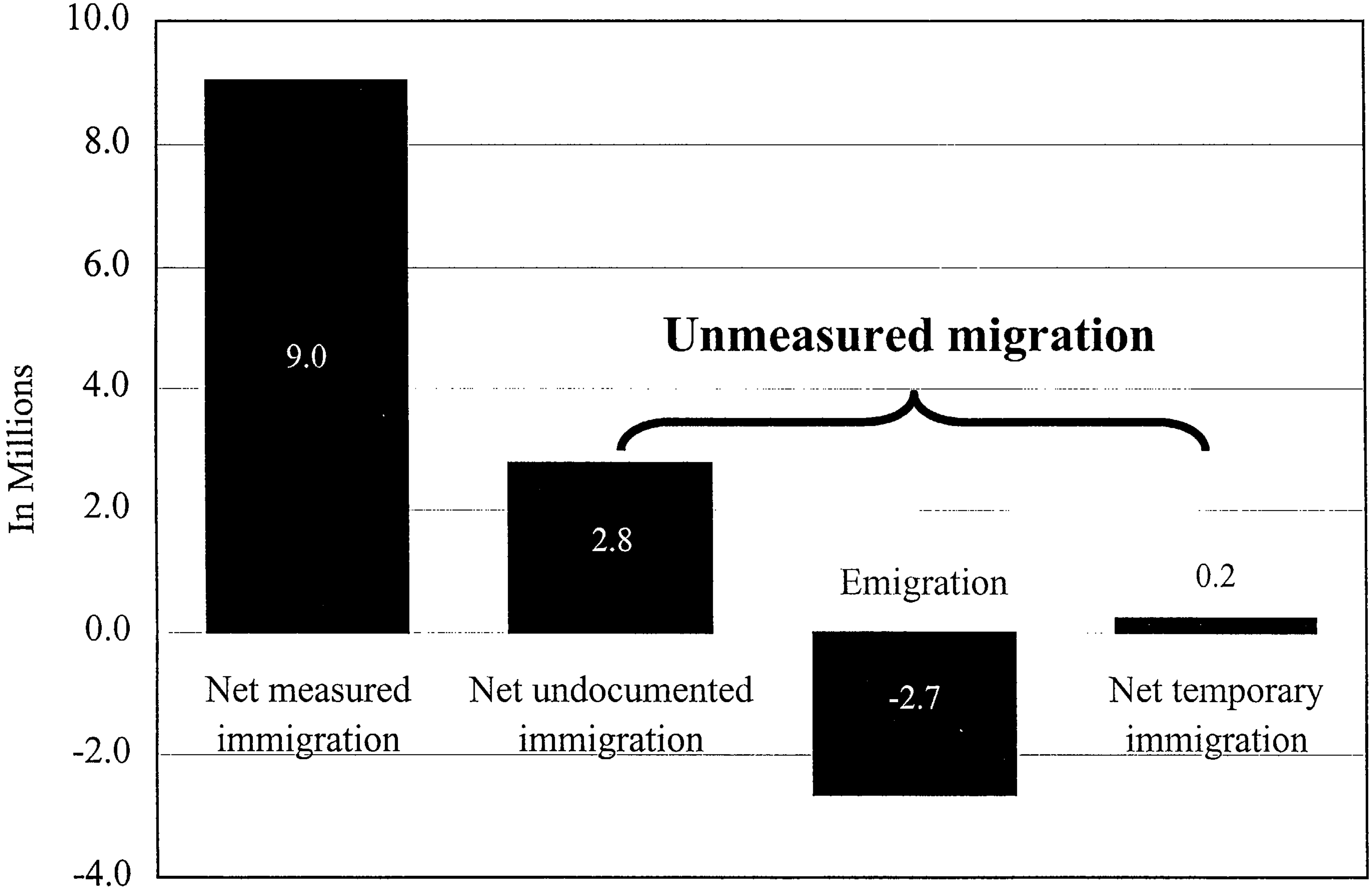


Population Accounting

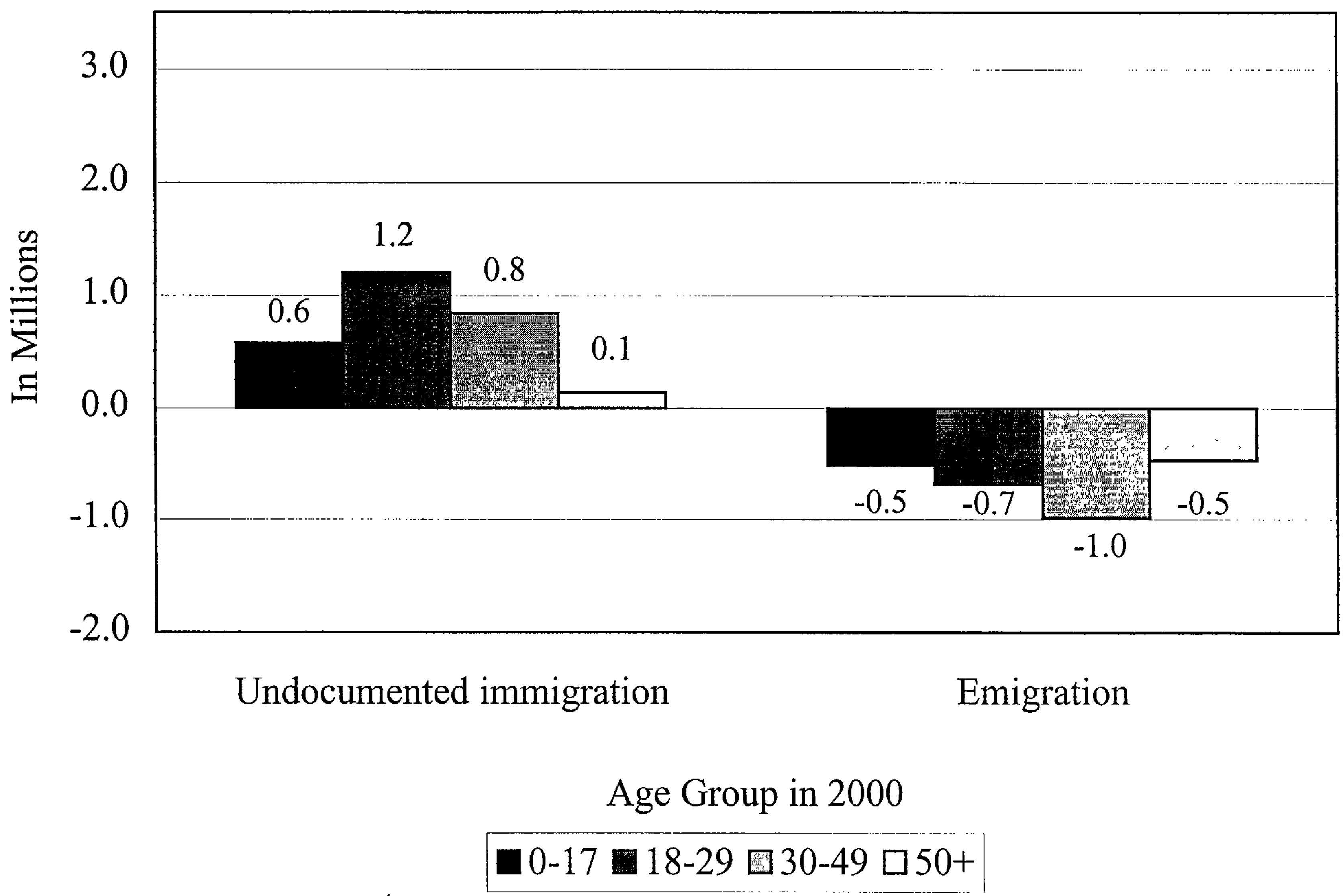
P1990 + PES + B - D + Migration + Unexplained = P2000+A.C.E.

Census 1990 (P1990)		248,700,000
PES (PES)		+ 4,000,000
Births (B)		+ 40,100,000
Deaths (D)		- 22,800,000
Net International Migration (Migration)		+ 9,400,000
Unexplained (Unexplained)		+ **5,300,000**
Census 2000 (P2000)	281,400,000	
A.C.E.	+ 3,300,000	
Adjusted Census 2000		284,7000,000

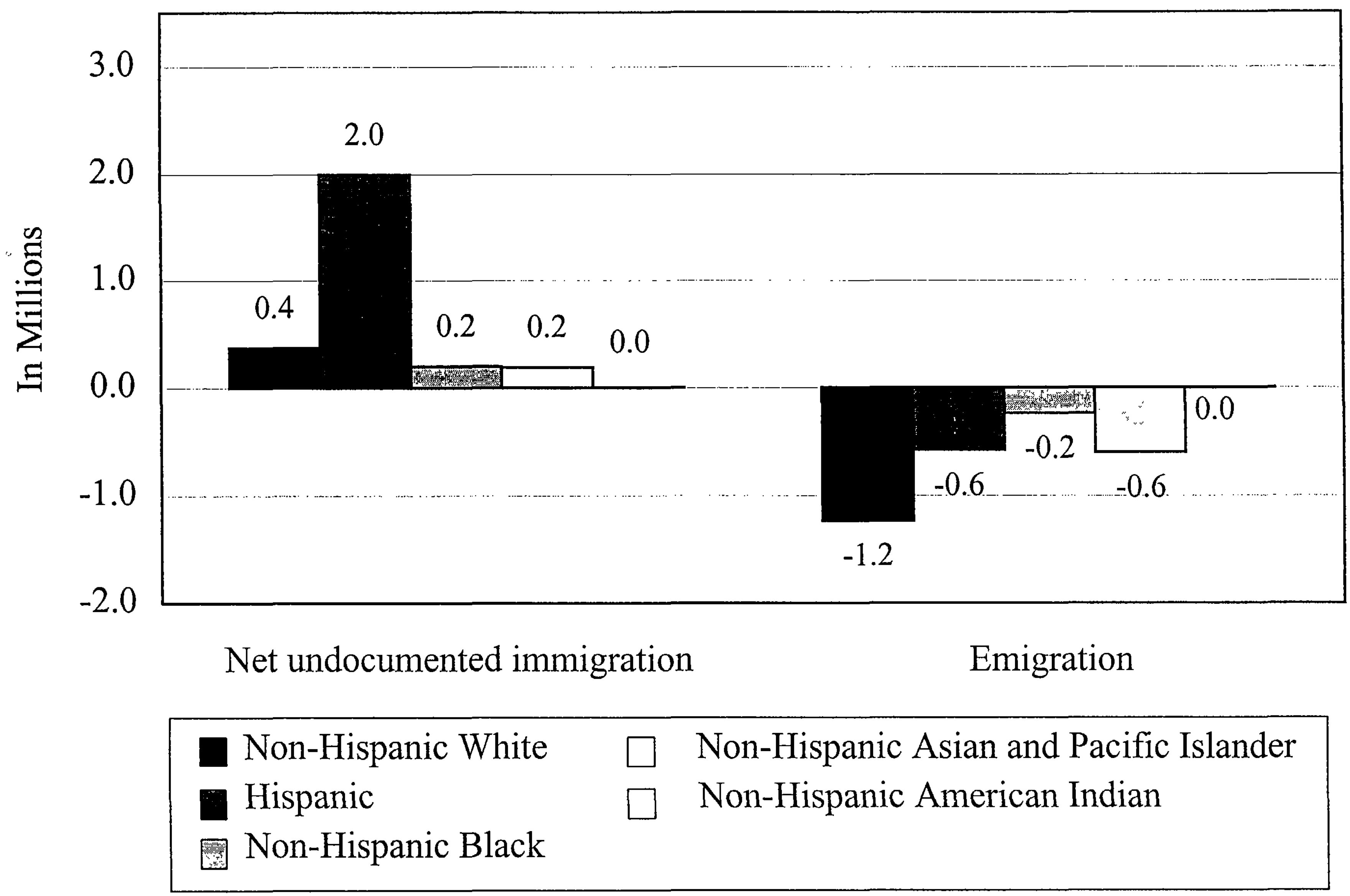
Current Components of Net International Migration: Measured vs. Unmeasured



Current Net Undocumented Immigration and Emigration by Age



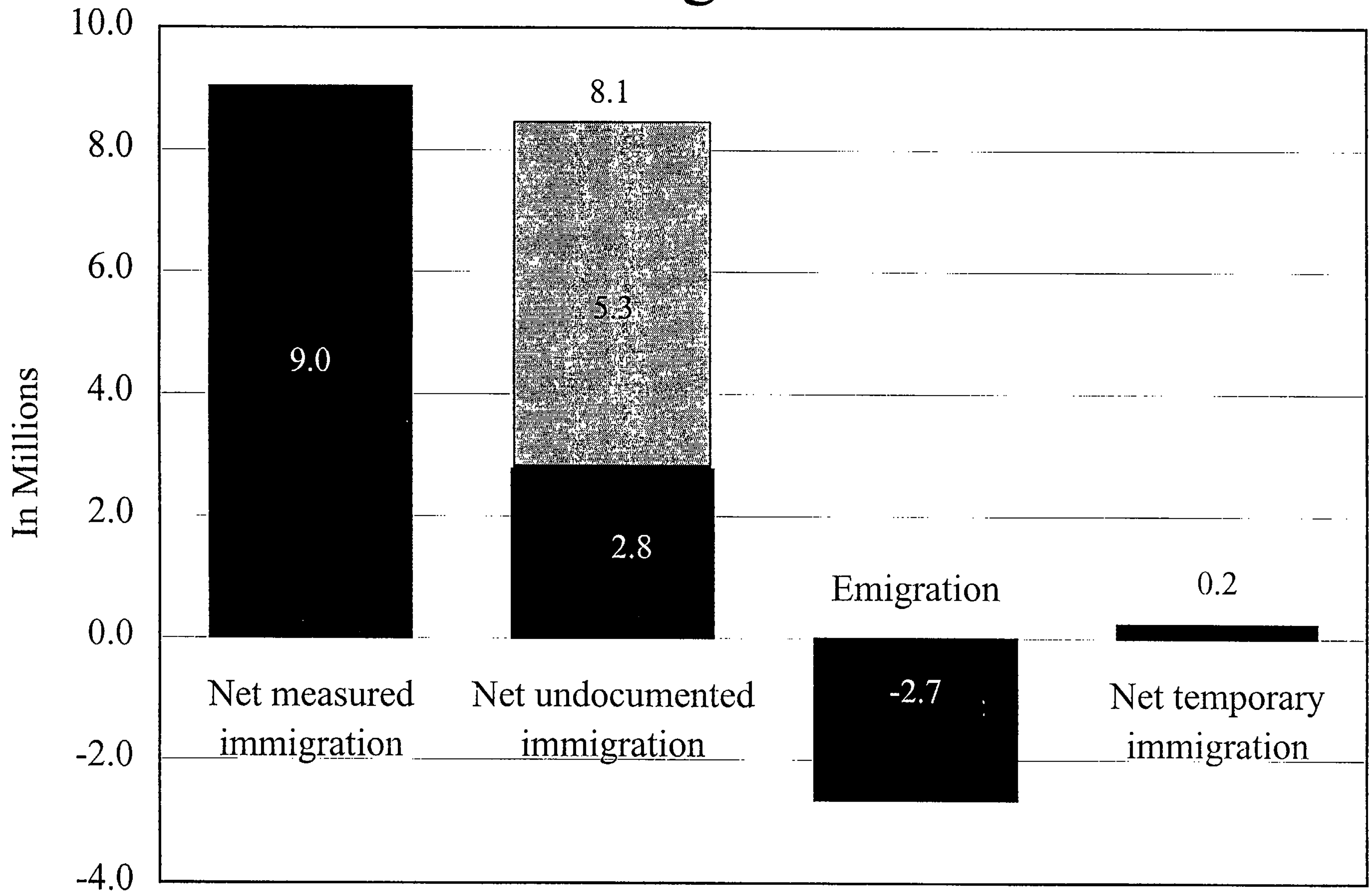
Current Net Undocumented Immigration and Emigration by Race and Hispanic Origin



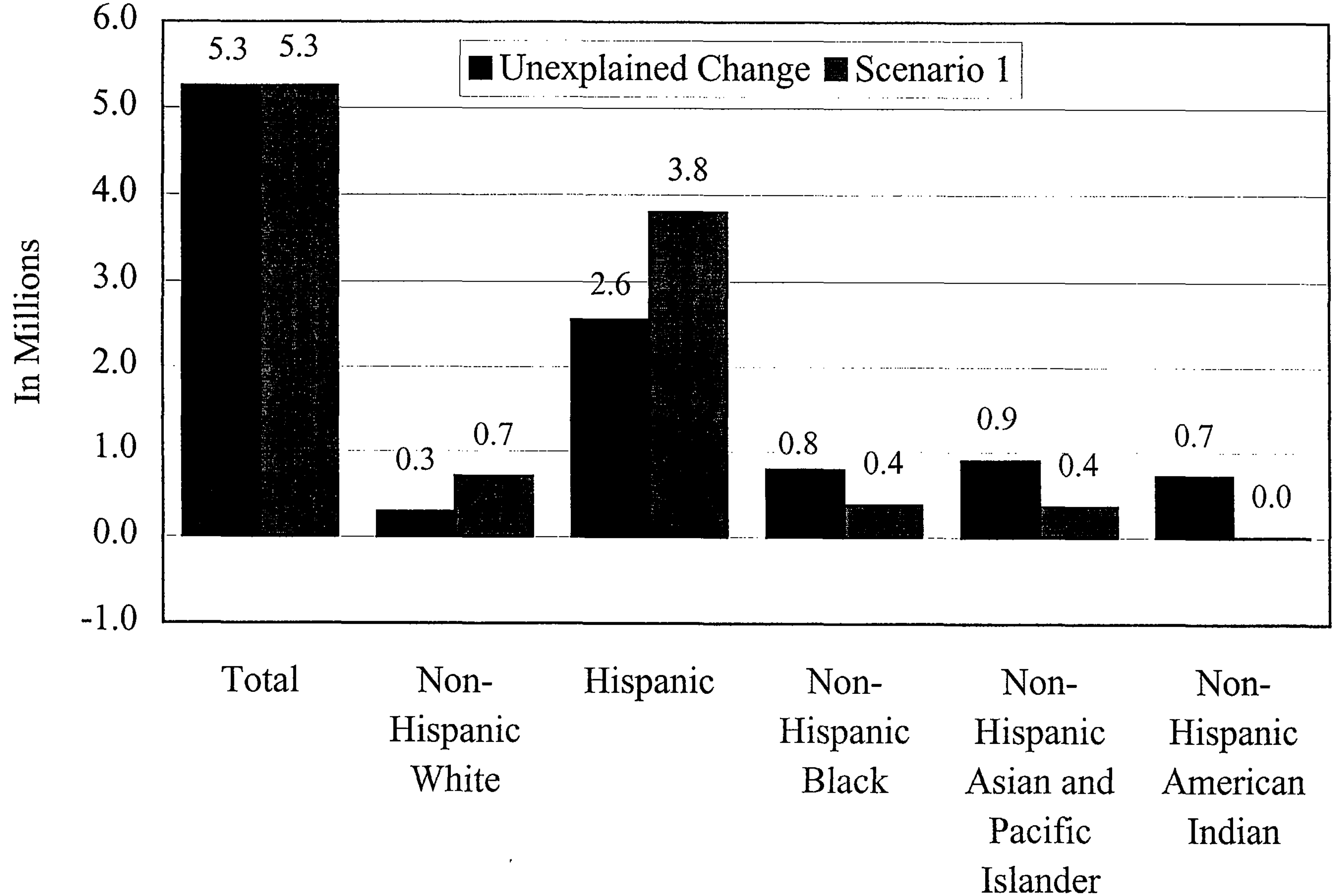
Alternate Demographic Scenarios Controlled to Census 2000 Adjusted

- Scenario 1: Increase undocumented
- Scenario 2: Zero emigration, remainder undocumented

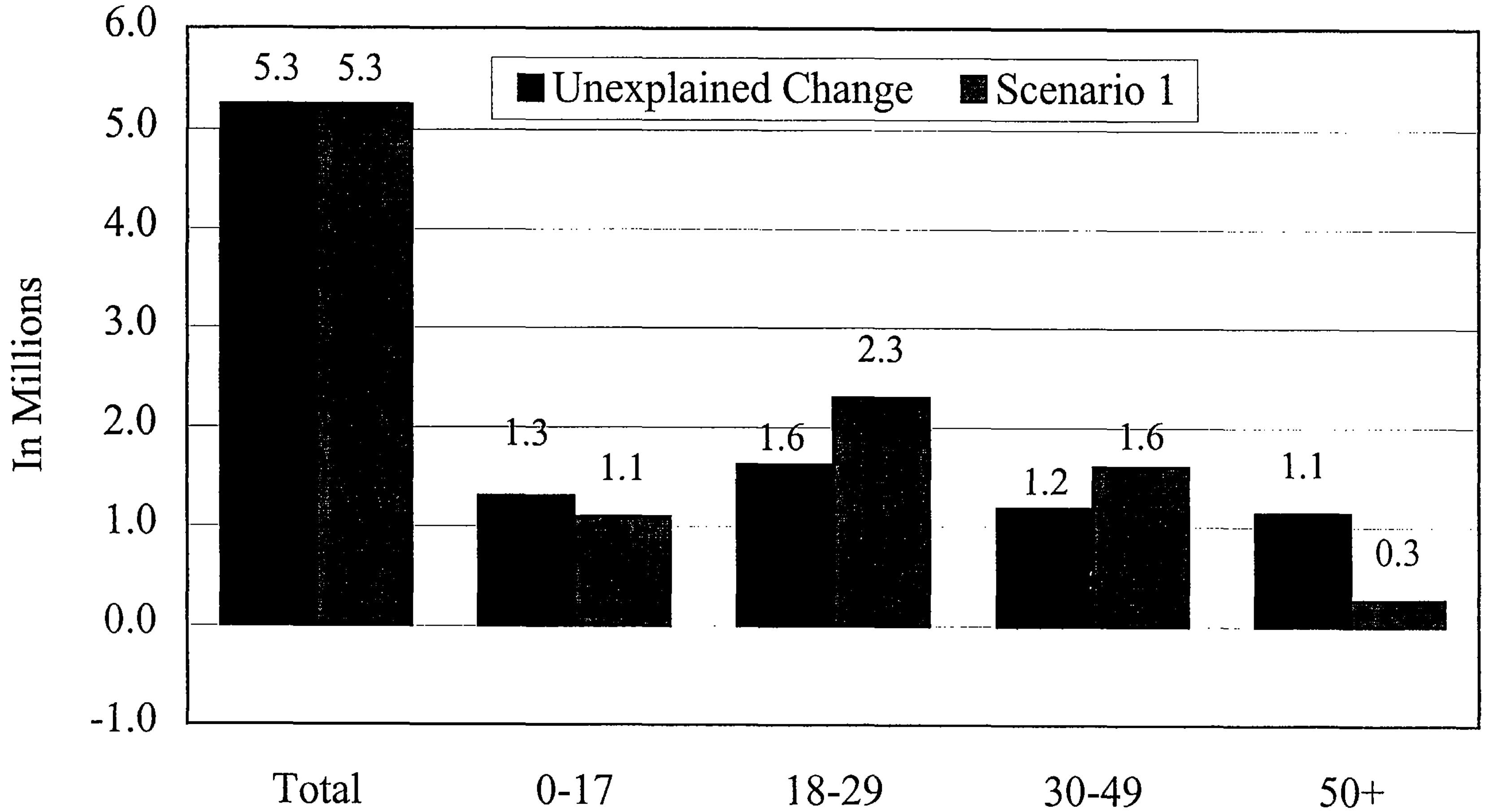
Scenario 1: Increase Undocumented Immigration



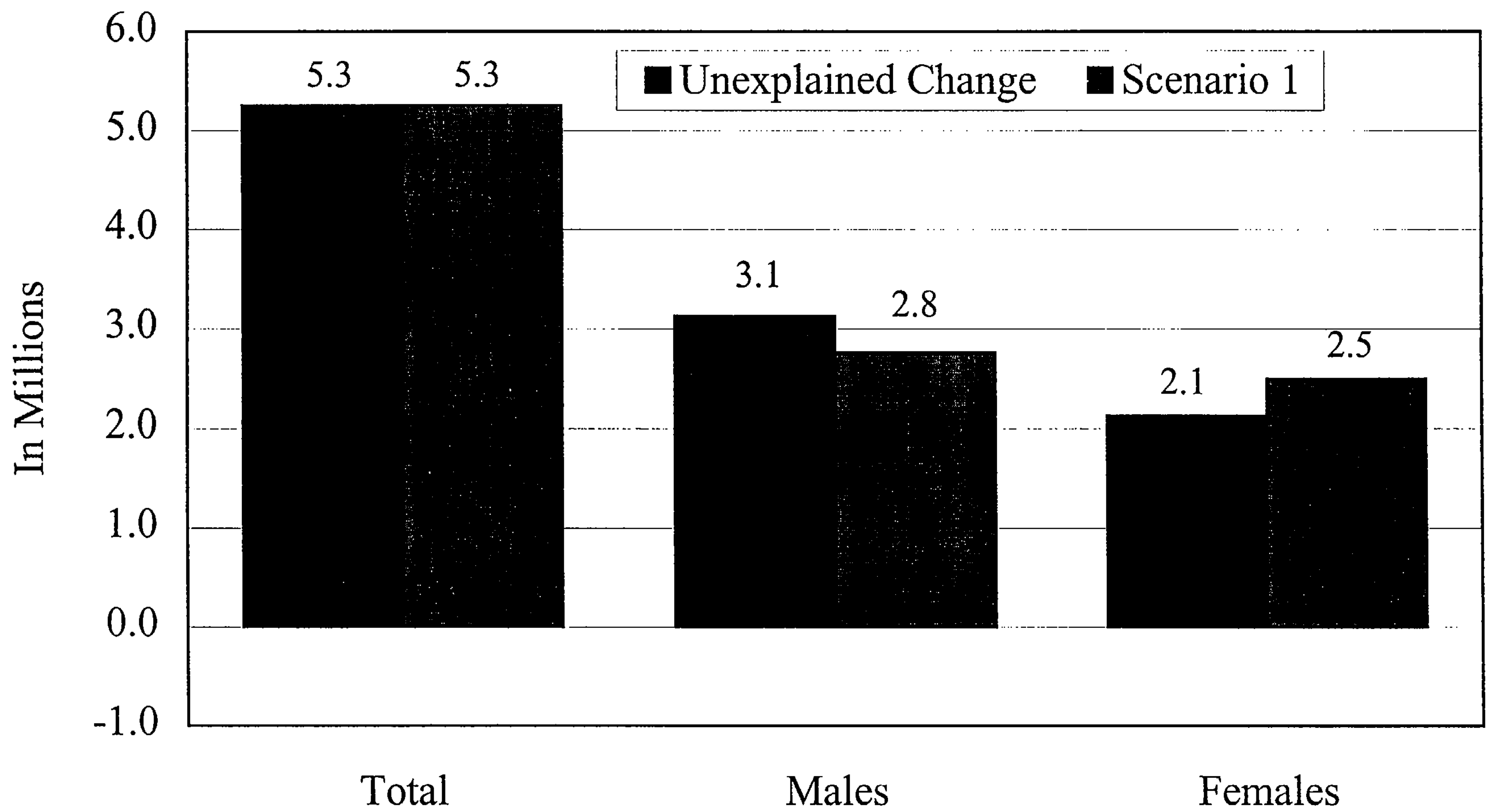
Comparison of Unexplained Change with the Results of Scenario 1 by Race and Hispanic Origin



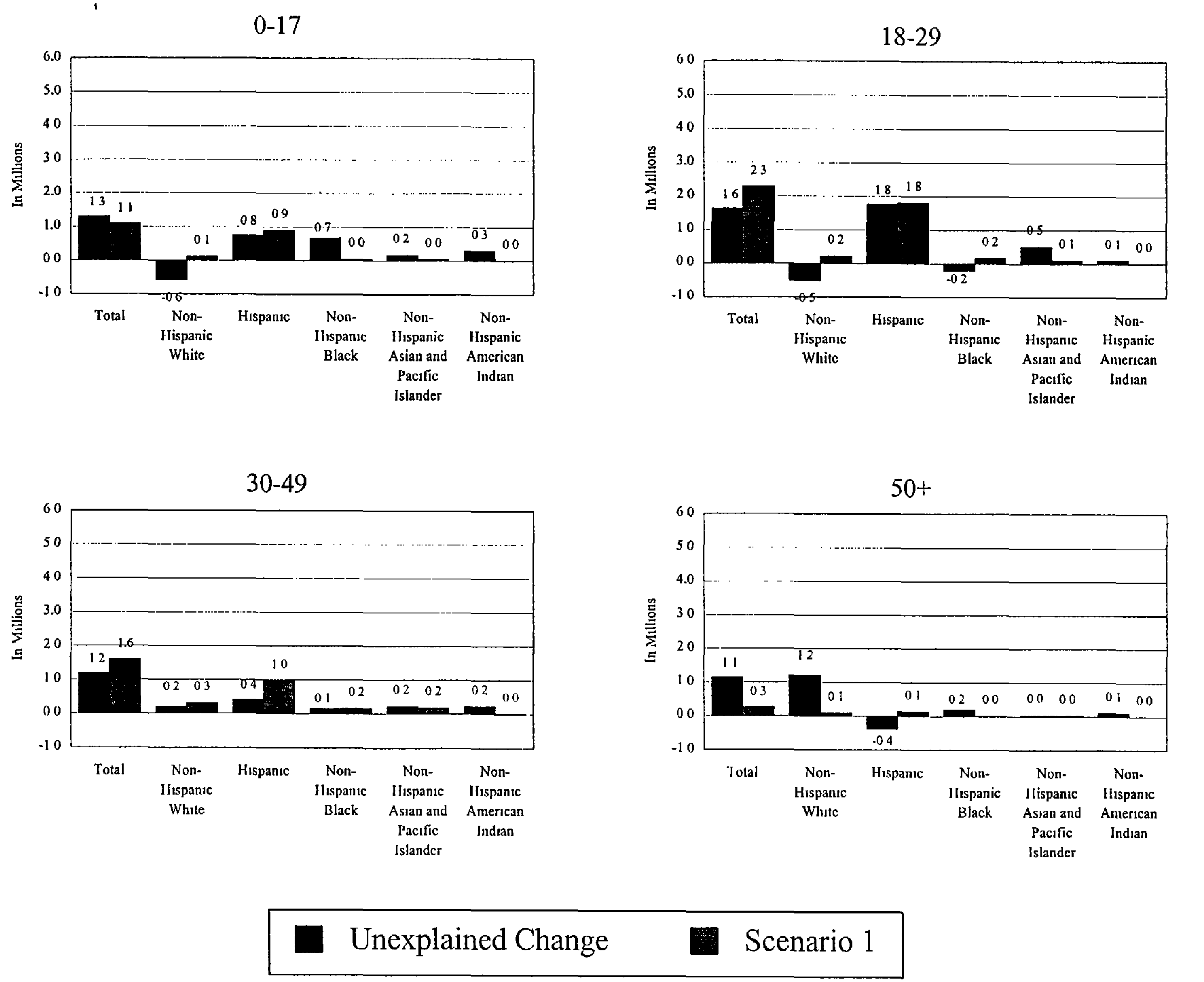
Comparison of Unexplained Change with the Results of Scenario 1 by Age



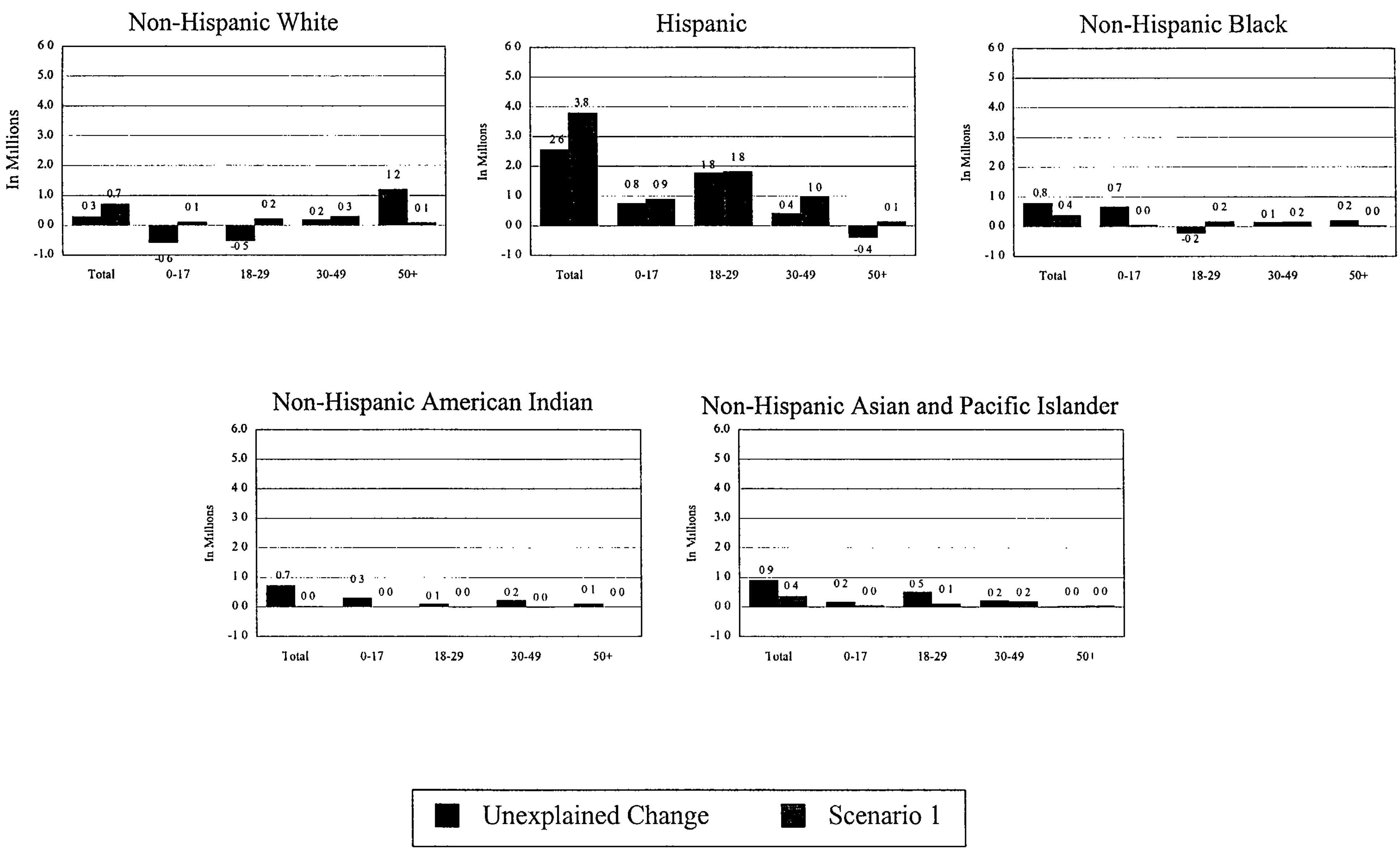
Comparison of Unexplained Change with the Results of Scenario 1 by Sex



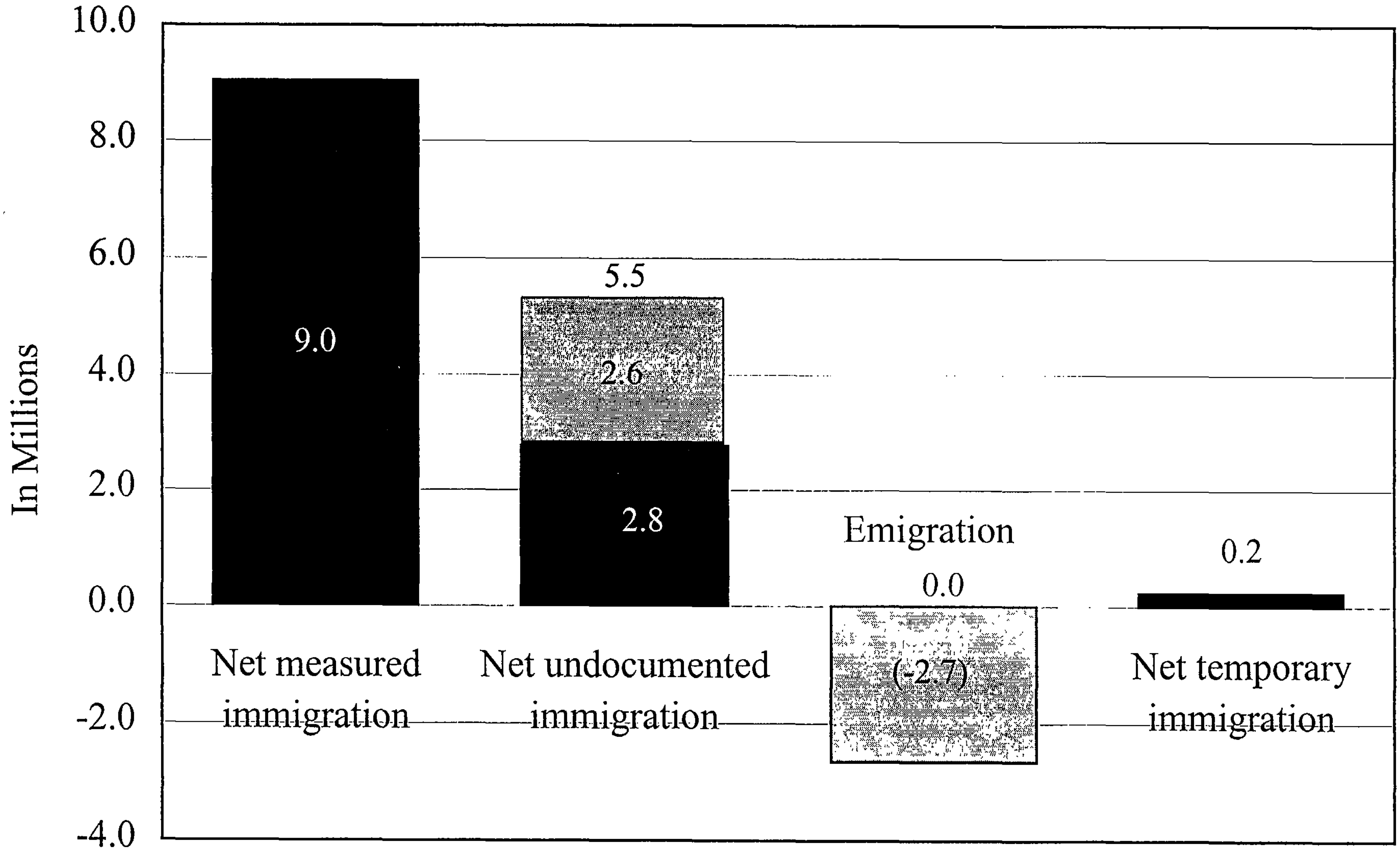
Scenario 1 by Race and Hispanic Origin for all Age Groups



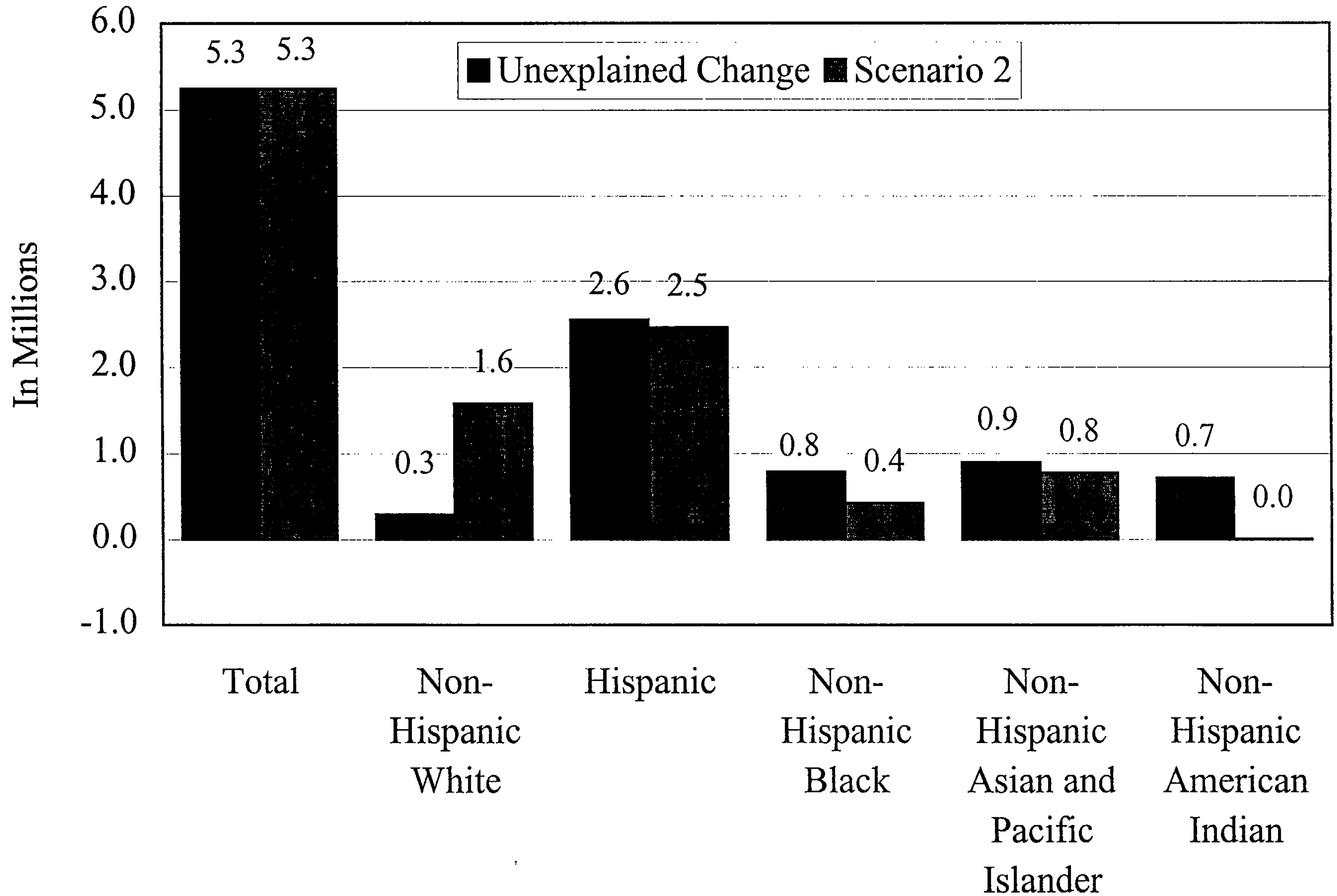
Scenario 1 by Age for all Race and Hispanic Origin Groups



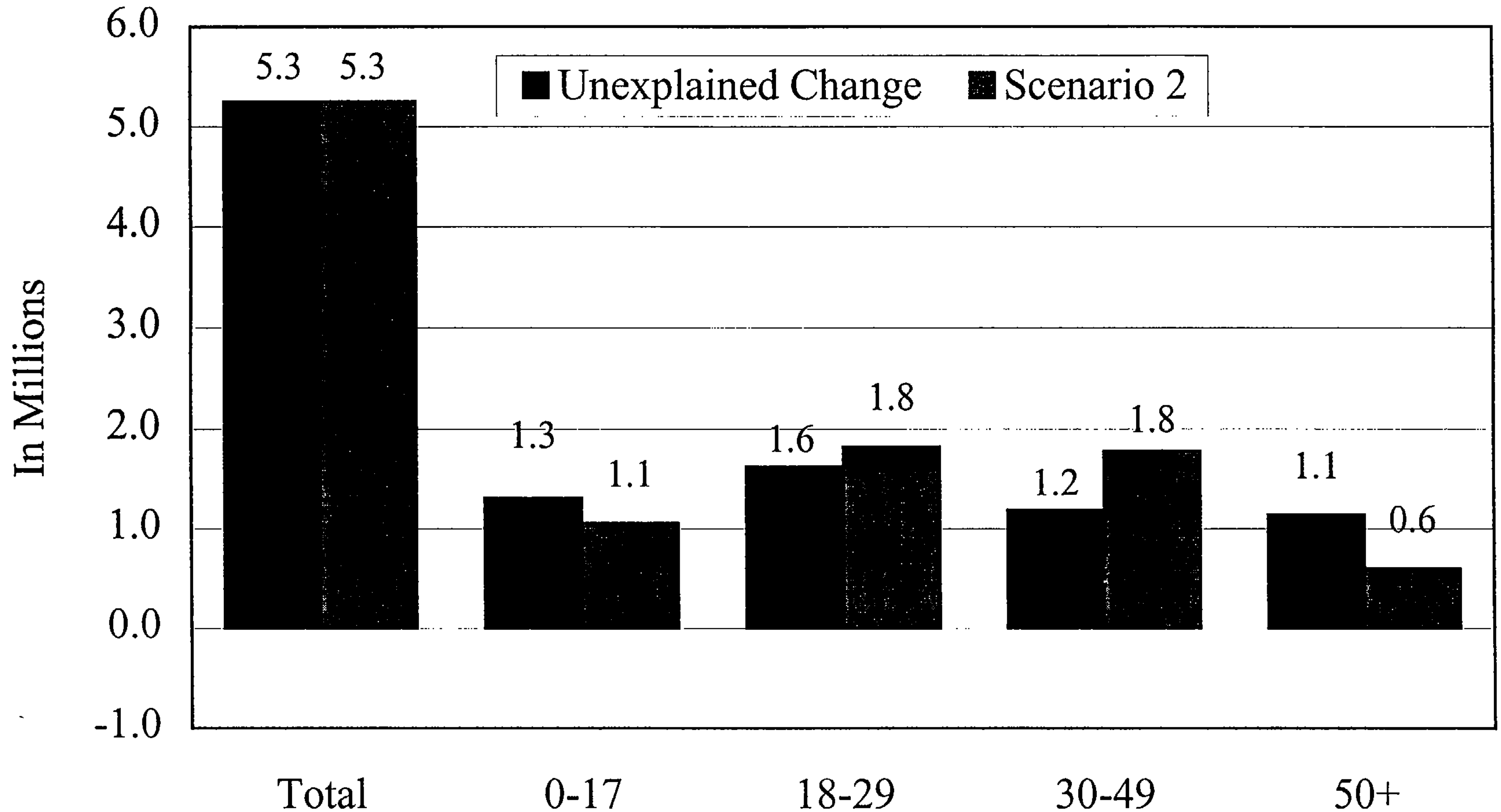
Scenario 2: Zero Emigration, Remainder Undocumented



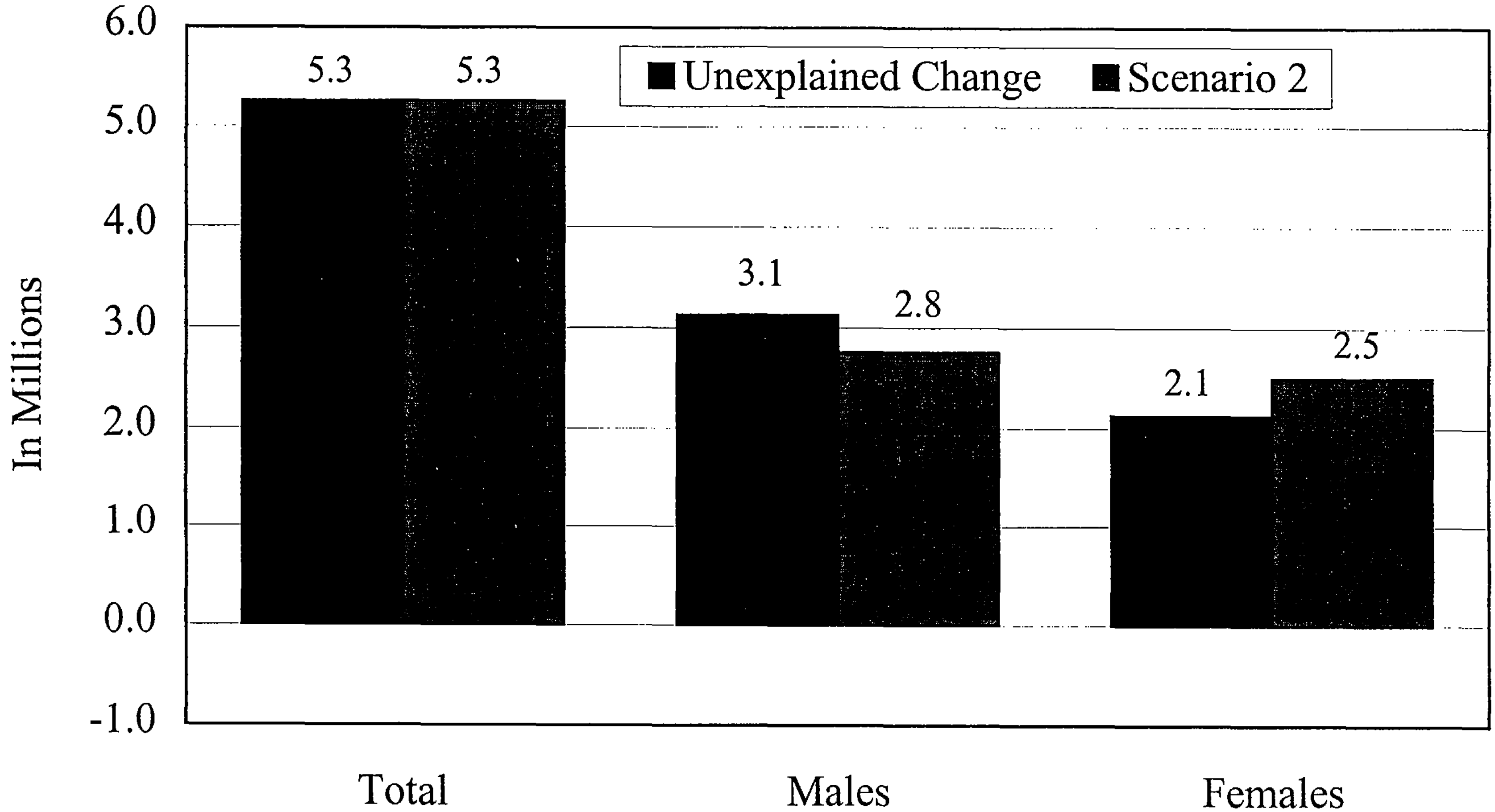
Comparison of Unexplained Change with the Results of Scenario 2 by Race and Hispanic Origin



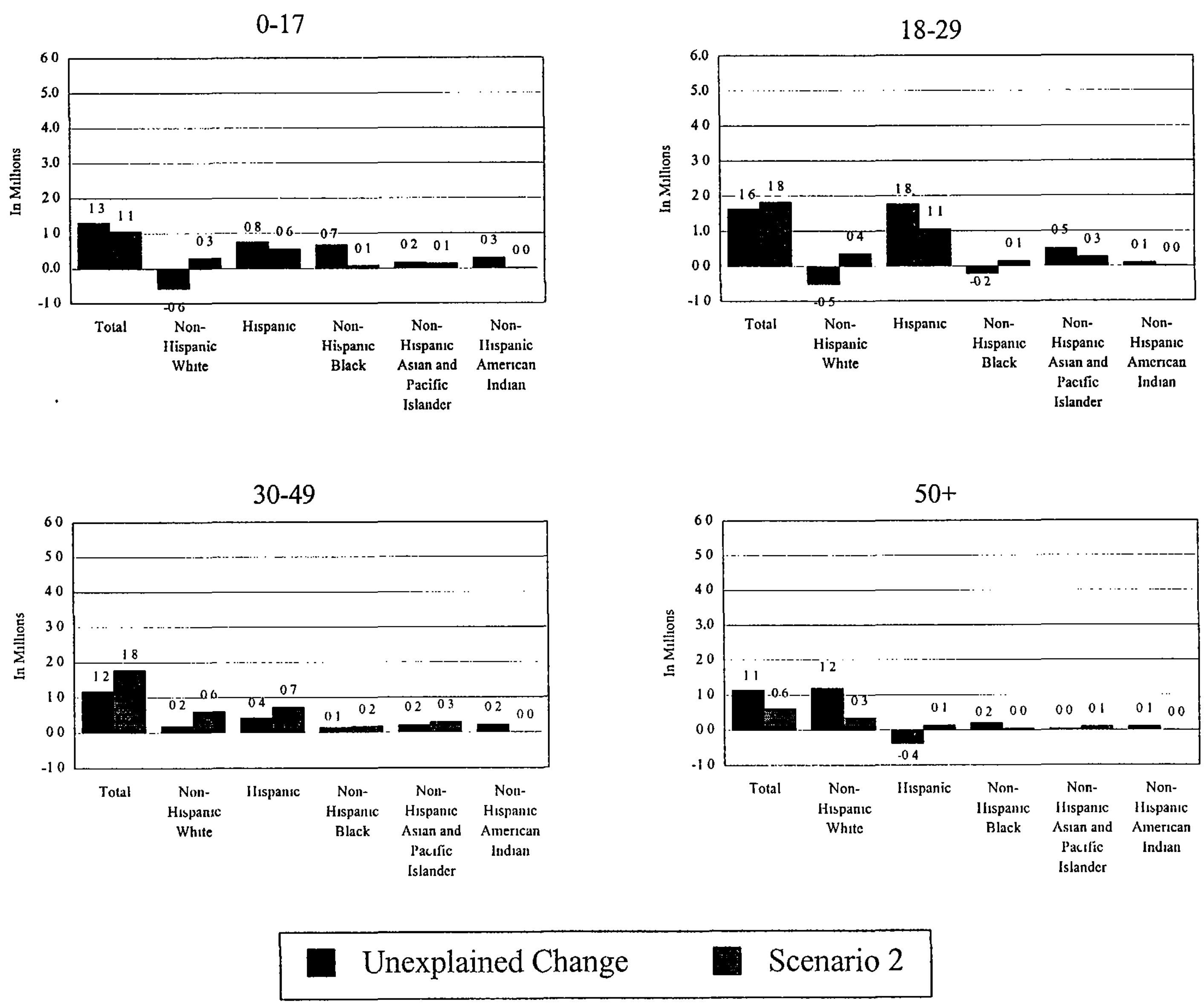
Comparison of Unexplained Change with the Results of Scenario 2 by Age



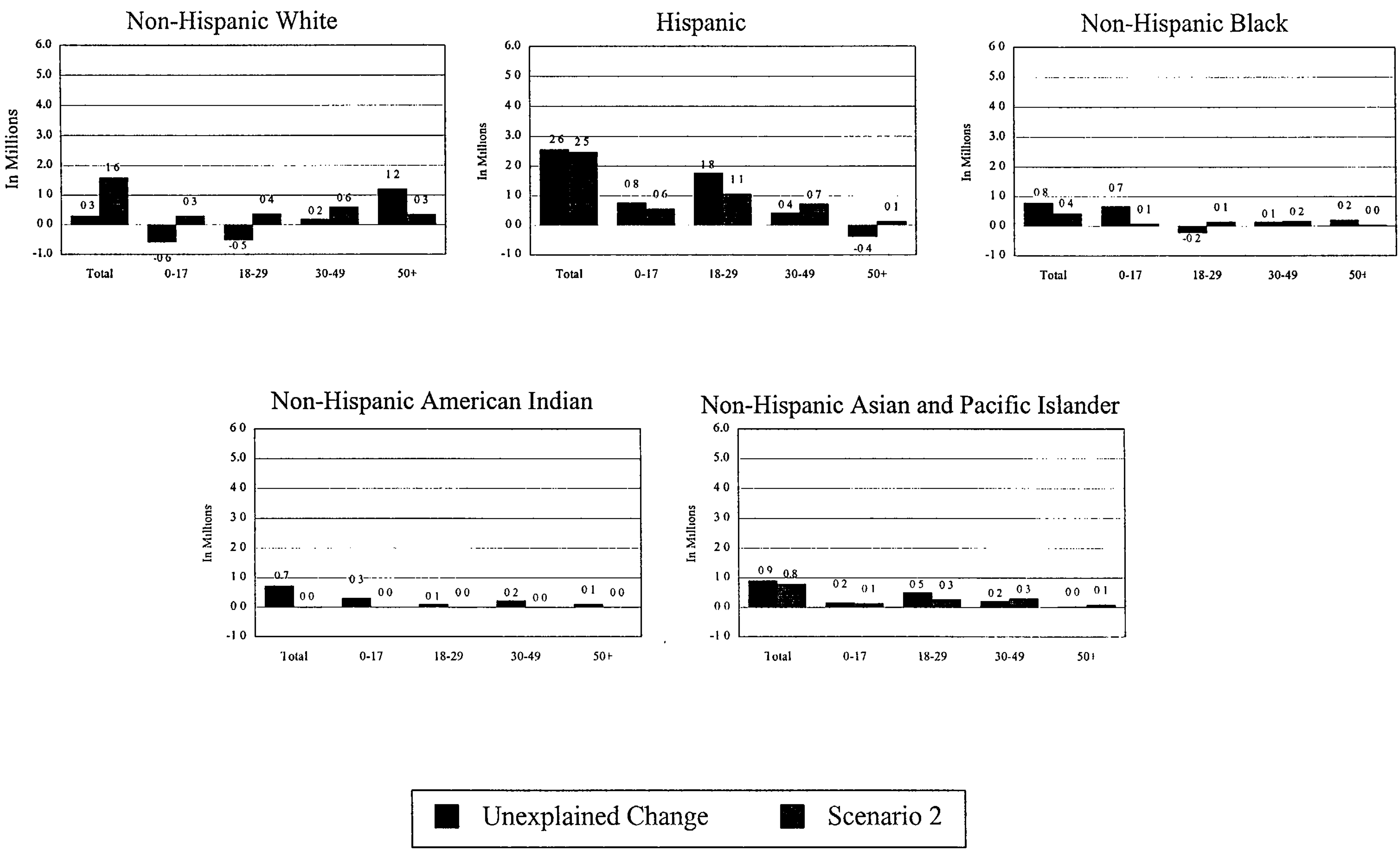
Comparison of Unexplained Change with the Results of Scenario 2 by Sex



Scenario 2 by Race and Hispanic Origin for all Age Groups



Scenario 2 by Age for all Race and Hispanic Origin Groups



Materials attached to these minutes were draft and preliminary material to inform the ESCAP Committee. The data and analysis contained in these documents are subject to revision and are not final. These materials report the results of research and analysis undertaken by Census Bureau staff. They have undergone a more limited review than official Census Bureau publications. Research results and conclusions expressed are those of the authors and do not necessarily indicate concurrence by the Census Bureau.

Implied Foreign-Born as a Percent of the Total Population by Alternative Scenorios

	2000 CPS			Base			Scenario 1			Scenario 2			Scenario 3		
Category	Total	Foreign Born	Percent of Population	Total	Foreign Born	Percent of Population	Total	Foreign Born	Percent of Population	Total	Foreign Born	Percent of Population	Total	Foreign Born	Percent of Population
US Total	274,087,000	28,379,000	10 35	279,467,263	28,586,991	10 23	284,721,906	33,841,635	11 89	284,655,308	33,361,602	11 72	283,558,757	32,438,469	11 44
Hispanic	32,804,000	12,841,000	39.14	33,790,648	12,342,855	36 53	37,590,046	16,142,253	42 94	36,246,773	14,766,935	40 74	36,079,337	14,613,573	40 50
Asian and Pacific Islander	10,925,000	6,706,000	61.38	10,721,484	6,651,452	62 04	11,084,498	7,014,466	63 28	11,501,825	7,424,428	64 55	11,212,455	7,139,091	63 67
Non-Hispanic	241,283,000	15,538,000	6.44	245,676,615	16,244,137	6 61	247,131,860	17,699,381	7 16	248,591,792	18,594,666	7.48	248,378,956	17,824,896	7 20

Materials attached to these minutes were draft and preliminary material to inform the ESCAP Committee. The data and analysis contained in these documents are subject to revision and are not final. These materials report the results of research and analysis undertaken by Census Bureau staff. They have undergone a more limited review than official Census Bureau publications. Research results and conclusions expressed are those of the authors and do not necessarily indicate concurrence by the Census Bureau.

The ESCAP has been presented a lot of data over the past several weeks. We have seen summary presentations from the detailed reports that Howard and staff are drafting. The purpose of this document is to present a summary of our findings to date, the issues we have identified, and questions that we have to address in order to reach a conclusion. I believe that in order to recommend the A.C.E. for the purposes of adjustment, we must be able to demonstrate that under robust assumptions regarding the total error model and the associated loss function analysis an improvement is achieved.

The majority of this document deals with concerns that we must address. However, it is useful to first briefly review the many positives regarding both the A.C.E. and Census 2000.

A.C.E. and Census 2000 Quality – Positives

<u>The A.C.E. was conducted as planned and with reasonably high quality</u>

The response rates for the interviewing phase were good, and in fact better than we expected (discussion to be provided).

Operations were completed on schedule.

Quality assurance operations were carried out as planned, and give preliminary evidence that the A.C.E. was in control (in the statistical sense). (Note, A.C.E. matching quality assurance is discussed below.)

Computer programs were throughly tested, there is a definite improvement from the level of testing in1990.

The sampling variances that we have seen indicate that we exceeded our design expectations. This is also a reflection on the quality of Census 2000, since clustered coverage errors would tend to increase these variances beyond our design expectations.

<u>Census 2000 also has some positives</u>

All major operations were completed as scheduled, and a few additional “clean-up” operations were implemented.

Quality assurance operations were conducted reasonably well.

All software was tested and verified.

Duplicates were removed

Both demographic analysis and the A.C.E. indicate that undercount levels were reduced from the 1990 census for the Black population. The A.C.E. also shows a reduction in undercount for the Hispanic population.

Demographic full count review presents anecdotal evidence that no serious clusters of Census 2000 coverage error exist.

Preliminary reviews of the E-sample findings (Howard will provide) indicate Census 2000 is reasonably comparable to 1990, further reenforcing the notion that no serious errors are present in the data.

There are some remaining questions regarding both Census 2000 and the A.C.E., however some limited conclusions can be drawn. Census 2000 was not conducted perfectly and without undercount or overcount. We most likely have the situation of both a good A.C.E. and a reasonably good census. There do not appear to be any fatal flaws in the data that would preclude adjustment, however, our job becomes more difficult in determining whether adjustment will improve redistricting data. As will be discussed in more detail below, we must carefully review and understand the sensitivity of the assumptions and parameters we using in our total error model and associated loss function analysis.

Concerns and Issues

The following seven sections address the concerns and issues that we need to resolve in order to provide our recommendation by March 1.

I. Demographic Analysis

Demographic analysis is telling us a different story than the A.C.E. While this may be due, in part to the changing complexity of our population, and the introduction of multiple response for race, it is important to understand these differences.

Concerns

> Given the relationship of demographic analysis, Census 2000, and the A.C.E. estimates, questions arise regarding whether the total error model is capturing all of the bias in the A.C.E. estimates.

Actions

1 John Long must address the differences between the A.C.E and DA. Our review of this discussion should focus on whether the A.C.E is overstating components of the population. Some questions that must be considered clearly include reexamination of DA assumptions on immigration and emigration, and whether the Census 2000 enumeration of the non-household population is contributing significantly to this difference.

2 Sex ratios are being used to produce measures of correlation bias. Given that the total error model and the loss function analysis are greatly influenced by correlation bias assumptions, it is essential that we review the assumptions underlying the assumption that these ratios are stable and predictive of coverage deficiencies.

3 The actual population levels produced by the total error model should be compared to demographic analysis.

II. Total Error Model

The total error model brings together all of the components of error that can be measured for the A.C.E. The components of error are parameters in the total error model. Several of these parameters can be estimated directly from Census 2000 data, but a number of the parameters must be derived from the 1990 results. The total error model is used to correct the A.C.E. for biases and thus is designed to produce a measure of “truth” that can be used to assess the accuracy of both the A.C.E. and Census 2000.

Truth can sometimes be elusive, and this is the case for the total error model. We refer to our measures of the truth as targets since we don't have unique estimates for all of the parameters in the total error model. This is because some of the parameters must be estimated based on models and assumptions that can vary. By using a range of targets as the basis of comparing the A.C.E and Census 2000, we can determine which situations favor the A.C.E. and which situations favor Census 2000. Situations are defined for our purposes by the methods and assumptions that we use in estimating the errors in the A.C.E -- the parameters in the total error model.

Given the limitations of our ability to measure nonsampling error (the parameters of the total error model) we can only produce direct measures of A.C.E. bias for 15 evaluation post-strata (groupings of the 448 A.C.E. post-strata). We must use models to apply these biases to the 448 A.C.E. post-strata for subsequent analyses of the A.C.E and Census 2000. For sensitivity purposes we use two models to accomplish this and compare the results. In addition, correlation bias has been shown to greatly influence the total error model, and thus further sensitivity is introduced into the total error model to examine the effect our estimates of correlation bias.

Loss functions are the tool that is used to compare the A.C.E. and Census 2000 with the targets that the total error model generates. Loss functions are discussed in the following section.

Concerns and Issues

Given the importance of the total error model in our analysis an critical concern is the degree to which the total error model reflects all of the errors in the A.C.E. We must have a thorough and extensive rationale for the measures produced by the total error model.

Actions

1 We need to do a sensitivity analysis by varying the assumptions underlying the parameters in the total error model. It will be very important to understand the sensitivity of the total error model to variations in the error parameters. This will allow for a documentation of the reasonableness of using 1990 parameters (e.g., if we only need a 10 percent improvement over 1990 to achieve large gains in accuracy for the A.C.E., we can take a strong position on relying on the total error model results). Sensitivity analysis will also allow for some assessment of the robustness of our implied assumption that the total error model reflects all of the error in the A.C.E.

2 Correlation bias has already been shown to have a large influence on the A.C.E. If no correlation bias is assumed, Census 2000 is shown to be more accurate in terms of both distributive and numeric accuracy for states and congressional districts. Clearly demographic analysis indicates that assuming no correlation bias is wrong, thus we need to determine, through sensitivity analysis, how much correlation bias we need to assume for a decisive result for the A.C.E. We also must examine additional models for correlation bias, including models

that do not include any correlation bias for owners, and models that assume correlation bias for the Hispanic population is similar to that for the Black population.

3 We have changed the treatment of movers since 1990. We must document how this effects the parameters we are using in the total error model.

III. Loss Function Analysis

As described above, loss function analysis is the tool that is used to compare the A.C.E. and Census 2000 with the targets generated from the total error model. The loss functions have been applied for states and congressional districts. We have used loss functions to assess both numeric and distributive accuracy. Thus far we have seen mixed results, depending on the assumptions we use for correlation bias.

Concerns and Issues

We have applied the loss functions for states and congressional districts. Clearly there are concerns about smaller areas, particularly since the redistricting data are specifically cited for state legislative districts.

We have four different loss functions, and we must be clear regarding those that we will rely on the most.

We are using the concept of expected loss instead of a hypothesis testing approach. We must be sure that all ESCAP members understand this concept.

Even though our goal is to select the data which are most accurate, we should know how much improvement we are expecting if we decide that adjustment is appropriate. This will be essential to put our recommendation in proper context. Currently we have not quantified the magnitude of the adjustment for congressional districts.

Actions

1 We should run the four (or a subset of the) loss functions for counties. This will allow for an assessment for something comparable to legislative districts. I would be concerned if we do not achieve numeric accuracy gains for the A.C.E.

2 We must document why we do not believe that blocks must be improved. This will be a question that we will be asked, so we may as well have the answer.

3 We must document our choices of the two loss functions that we favor.

4 We must have a discussion of the concept of expected loss.

5 We must develop a. quantitative measure of the change we are introducing to congressional districts as a result of adjustment. I would suggest as a measure the relative gains in the range of within state shares for congressional districts. (Example to be discussed)

IV. Synthetic Error

The A.C.E. adjustment is applied by calculating a coverage correction factor for each post-stratum. This factor is the ratio of the DSE to the Census 2000 count for the post-stratum. In effect we multiply each block by the coverage correction factors for the post-strata that the block includes, and then summing the blocks to larger areas of interest. It must be noted that this design implies that we are only correcting for systematic biases, and not local Census 2000 errors.

Synthetic estimation is based on the assumption that net coverage does not vary within the A.C.E post-strata. Failures of this assumption are referred to as synthetic error. Ideally synthetic error would be measured by simply computing A.C.E. direct measures of the coverage error for any set of sub-national areas of interest and comparing these to the synthetic estimate. Unfortunately, we do not have enough A.C.E. sample to accomplish this method of measurement. Thus we must rely on "artificial population" analysis. That is, we construct populations that have variables that we believe are distributed similar to coverage error.

Our preliminary analysis of artificial populations for states and congressional districts has indicated that we most likely do not have a situation where the net coverage is distributed in such a fashion that the synthetic estimate will result in improving only a few areas at the expense of the rest. We have also seen that within these populations we cannot ignore the effect of synthetic error as it is large in comparison with the errors in the DSE. This is very important because we do not include a measure of synthetic error in the total error model.

Concerns

The total error model does not include a measure of synthetic error. Therefore, the finding that synthetic error is large relative to the DSE error must be assessed.

We have four artificial populations. Concerns will be expressed that these do not reflect the distribution of coverage error. In addition, one of the populations gives rather extreme results for synthetic error.

Actions

1 In 1990 Fay and Thompson (mostly Fay) conducted an analysis showing that the loss function comparisons were conservative in favoring adjustment for all but one of the artificial populations we used when synthetic error is considered.. This analysis must be repeated for the artificial populations we are now using. If we find that the current artificial populations do not have this same conservative or at least neutral feature, we must then discuss how we assess the loss function analysis – we must place some conservative requirements on the level of improvement we must see to conclude that adjustment is warranted.

2 We must review the distributions of artificial population coverage errors. I'd suggest that we look at the post-stratum groups cross classified by region to determine if we have a consistent distribution of net coverage error.

3 We must document why we believe that population 3 is an outlier.

V. Targeted Extended Search (TES) and Balancing

It is important to use the same area for defining E-sample correct enumerations and P-sample matches. If this is not the case, then balancing error results, leading to either too many matches relative to correct enumerations or too few. We have seen the results of the targeted extended search and have found more matches than correct enumerations in surrounding blocks. In the absence of P-sample geocoding error this would be problematic.

Concerns

Is balancing error present in the DSE estimates? In 1990 we concluded that we did not have a. problem with balancing error with similar findings. However, we have refined our procedures for 2000 and must reexamine this issue.

We have an evaluation in progress that may shed some light on this, and our preliminary findings indicate that we may have incorrectly coded some E-sample cases as correct enumerations in the A.C.E. block cluster when they should have been coded in the surrounding block. This explains some of the discrepancy between the P-sample matches and the E-sample correct enumerations, however the evaluation also shows that some E-sample correct enumerations should have been classified as geocoding error. We are questioning whether the total error model includes measures of this error.

Bob Fay has been studying this situation, and has raised concerns that the TES estimates of duplicates are too low.

Actions

1 We should examine the A.C.E. quality assurance records to see if we can get a measure of P-sample geocoding error. This may explain the discrepancy between the P-sample matches and E-sample correct enumerations.

2 We will examine the parameters in the total error model to determine whether the error that we identified above is, in fact, included.

3 We must have a discussion on Fay's concern regarding the TES estimate of duplicates.

4 If we cannot resolve the issue of balancing, we must discuss how we build conservatism into our assessment of the loss function analysis.

VI. Late Adds and Whole Person Imputations (IIs)

We have a process for including late adds and IIs into estimation of coverage. However, the number of late adds and IIs have increased significantly for Census 2000.

Concerns

Given the larger than expected number of late adds and the attention that many stakeholders are expressing regarding these, it is necessary to document the theoretical basis for our current treatment.

There are also concerns that the late adds and IIs may be symptomatic of problems with the uniformity of net coverage within post-strata in the synthetic assumptions underlying the adjustment process.

Actions

1 We need to document the assumptions underlying our treatment of late adds and IIs.

2 We should review tabulations of the late adds and IIs within post-strata. I'd suggest post-stratum groups by region. If we observe serious deviations, we will take the actions discussed above for synthetic error.

VII. Other Actions Identified

1 We need to review the variances for counties smaller than 100,000 persons relative to 1990. This will demonstrate the accuracy of our results relative to 1990 for areas such as legislative districts.

2 We have examined and had discussions about our procedures and results for missing data, and we did not identify any serious problems. We have to be sure that this is documented.

3 Negative adjustments will be made if we decide in favor of adjustment. We need to have documentation prepared to address concerns that many critics have expressed.

4 A followup discussion is required to cleanup the situation regarding quality assurance results for A.C.E. matching. The previous discussion was somewhat confusing. This should be a short presentation that will demonstrate we have gained in the accuracy of matching relative to 1990.

ESCAP MEETING NO. 41 - 02/19/01

MINUTES

Minutes of the Executive Steering Committee on Accuracy and Coverage Evaluation (A.C.E.) Policy (ESCAP) Meeting # 41
February 19, 2001

Prepared by: Annette Quinlan and Maria Urrutia

The forty-first meeting of the Executive Steering Committee on Accuracy and Coverage Evaluation Policy was held on February 19, 2001 at 10:00.

The agenda for the meeting was to identify outstanding issues for the Committee's upcoming deliberations, and to discuss revised Demographic Analysis estimates.

Committee Attendees:

Nancy Potok
Paula Schneider
Cynthia Clark
Nancy Gordon
John Thompson
Jay Waite
Bob Fay
Howard Hogan
Ruth Ann Killion
John Long
Carol VanHorn

Deputy Director/Acting Director:
William Barron

Other Attendees:

Marvin Raines
Nick Birnbaum
Kathleen Styles
Maria Urrutia

I. Outstanding Issues

John Thompson distributed the attached document which presents a summary of the findings to date, the issues that have been identified, and questions that we have to address in order to reach a conclusion. The document contains talking points from which John led the discussion.

II. Demographic Analysis

John Long presented a comparison of the adjusted census in 1990 to the adjusted census in 2000. The Committee noticed some features that were hard to explain by the traditional emigration and immigration models. The next step is to analyze differences between the unadjusted census in 1990 and the unadjusted census in 2000. The Committee should look at these analyses when they are completed and see which differences are more plausible.

III. Next Meeting

The agenda for the next meeting, to be held on February 20, 2001, is to discuss different scenarios to account for the difference between the A.C.E. and Demographic Analysis, an analysis of Census 2000 Group Quarters and the independent estimates for Group Quarters, late census data, difference between movers in 1990 and 2000, and explanation of expected loss.

Review status of this document: Sent to Urrutia 2/20/01.

ESCAP MEETING NO. 42 - 02/20/01

AGENDA

There was no agenda developed or used for the February 20, 2001 meeting.

ESCAP MEETING NO. 42 - 02/20/01

HANDOUTS

Materials attached to these minutes were draft and preliminary material to inform the ESCAP Committee. The data and analysis contained in these documents are subject to revision and are not final. These materials report the results of research and analysis undertaken by Census Bureau staff. They have undergone a more limited review than official Census Bureau publications. Research results and conclusions expressed are those of the authors and do not necessarily indicate concurrence by the Census Bureau.

Proposed Demographic Activities for Decision Process

1. Run 3 alternative scenarios to fit unexplained difference between adjusted 1990 (PES) and adjusted 2000 (ACE).

   A. Scenario 1 - unexplained difference in net undocumented immigration.

   B. Scenario 2 - no emigration with remaining unexplained difference in net undocumented immigration.

   C. Scenario 3 - 50% reduction in emigration with remaining unexplained difference in net undocumented immigration.

   D. Determine what changes in components (or 1990 PES results) would give a plausible match to 2000 ACE.

2. Run 2 alternative scenarios to fit unexplained difference between adjusted 1990 (PES) and unadjusted Census 2000.

   A. Scenario A - new unexplained difference in net undocumented immigration

   B. Scenario B - 25% reduction in emigration with remainder of new unexplained difference in net undocumented immigration.

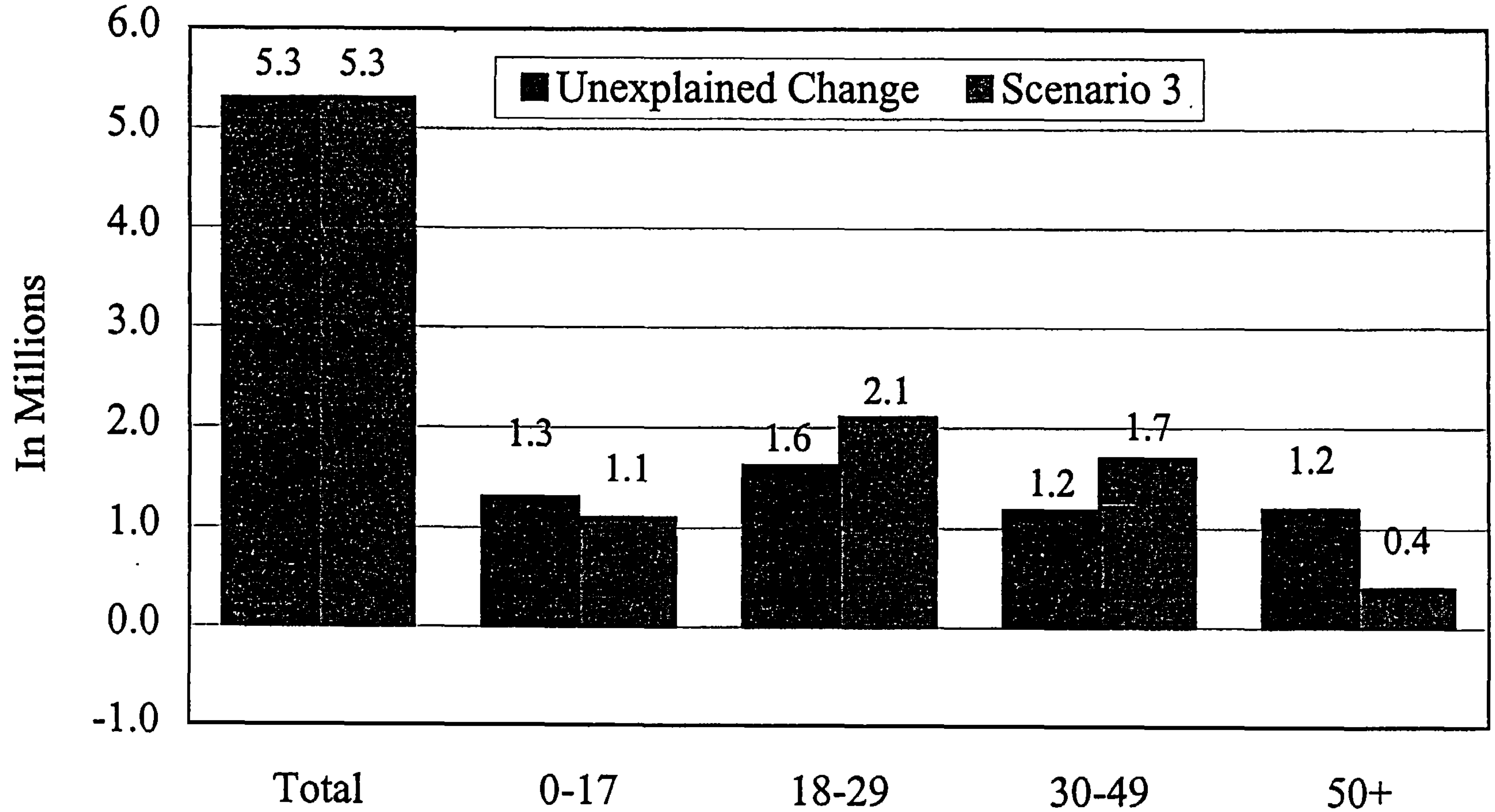
OR 2. Run the 3 alternative scenarios listed in 1 to fit the unexplained difference between 1990 unadjusted and unadjusted Census 2000.

3. Run a revised set of traditional demographic analysis with a doubling of undocumented immigration and no change in emigration. Compare with adjusted and unadjusted 2000 results.

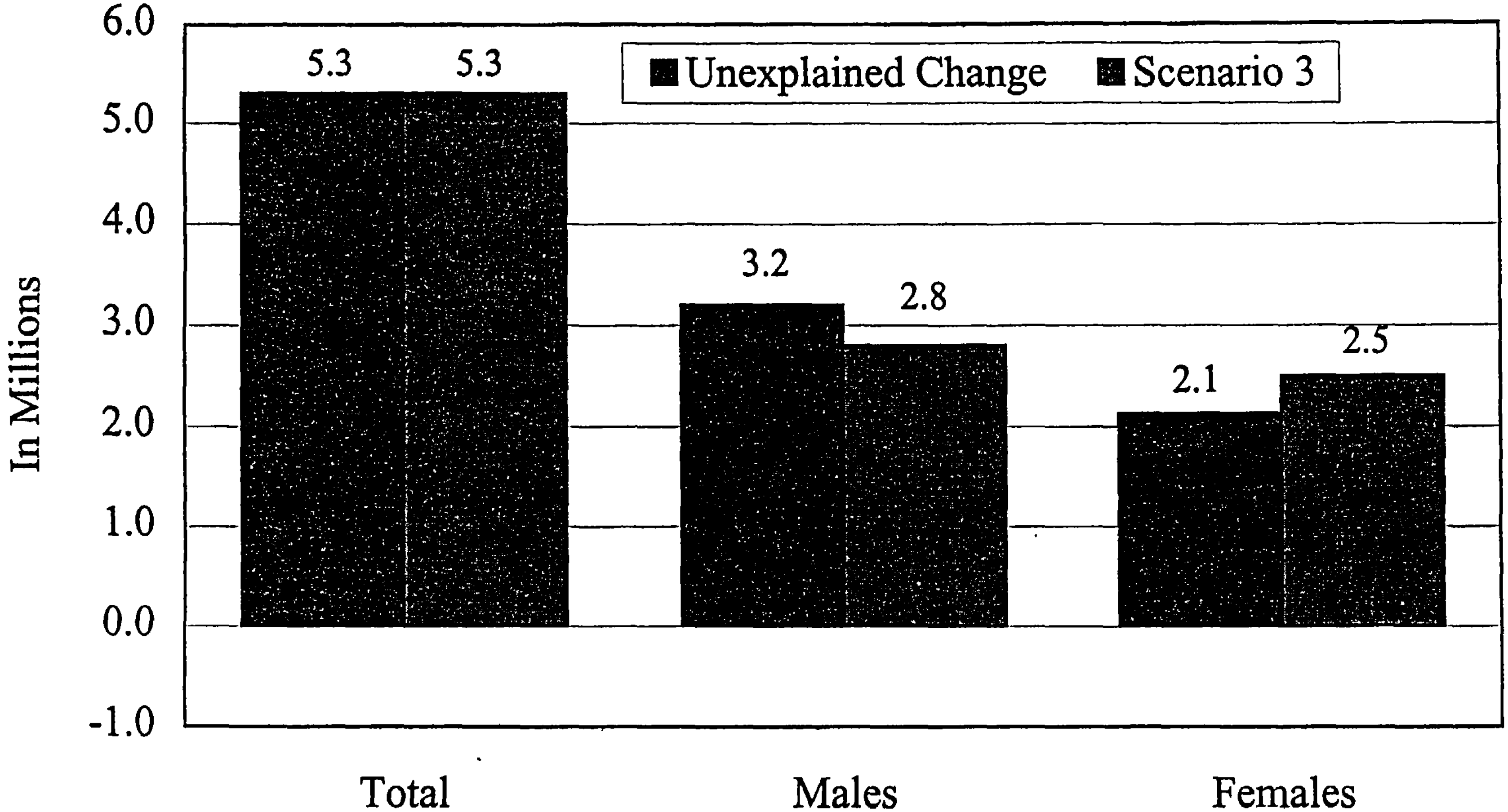
4. Compare the foreign-born percentages of all the above scenarios with the 2000 CPS foreign-born results. Also rerun the 2000 CPS without the adjustment to population estimates controls.

5. Present an evaluation of the coverage of the group quarters population.

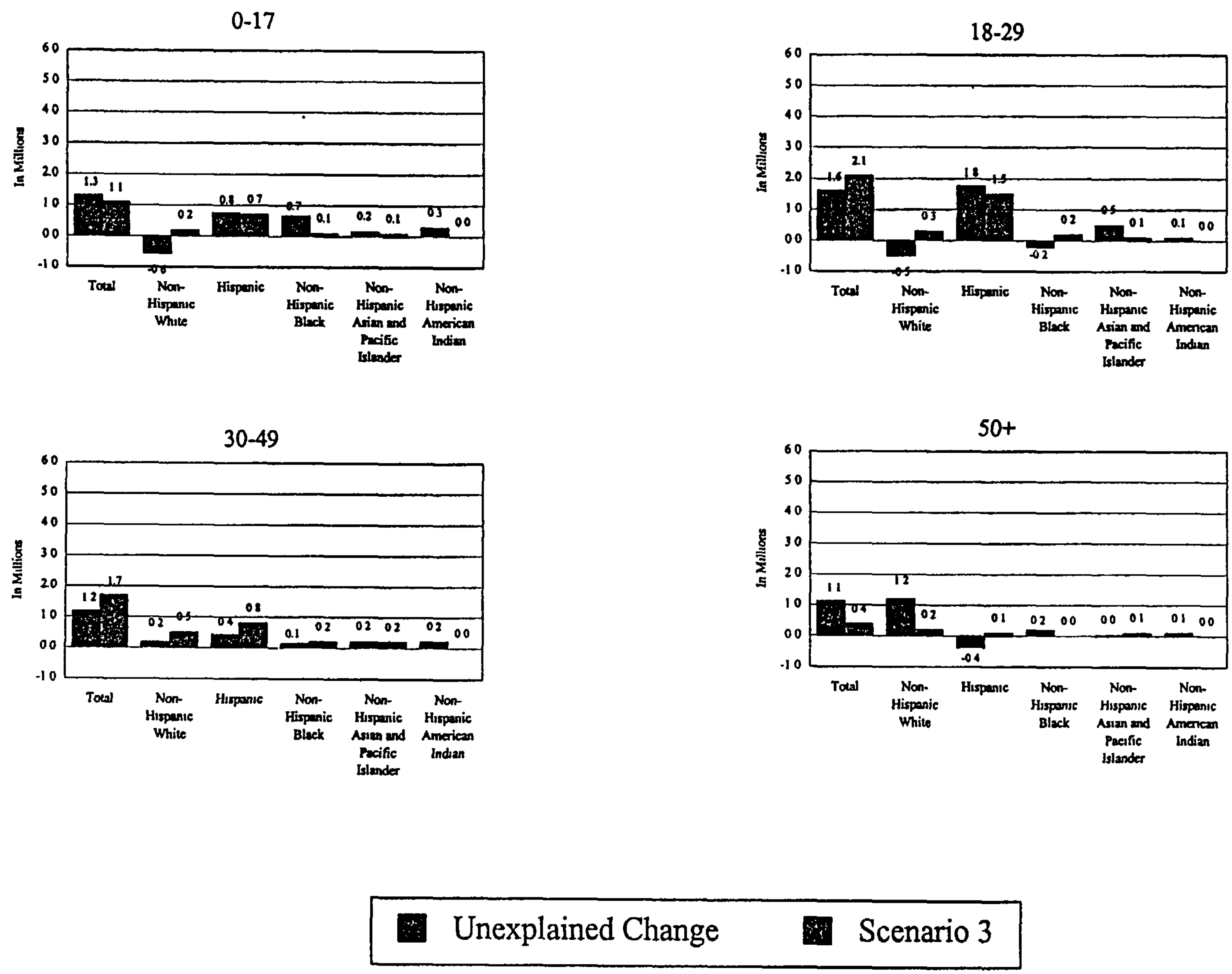
Comparison of Unexplained Change with the Results of Scenario 3 by Age



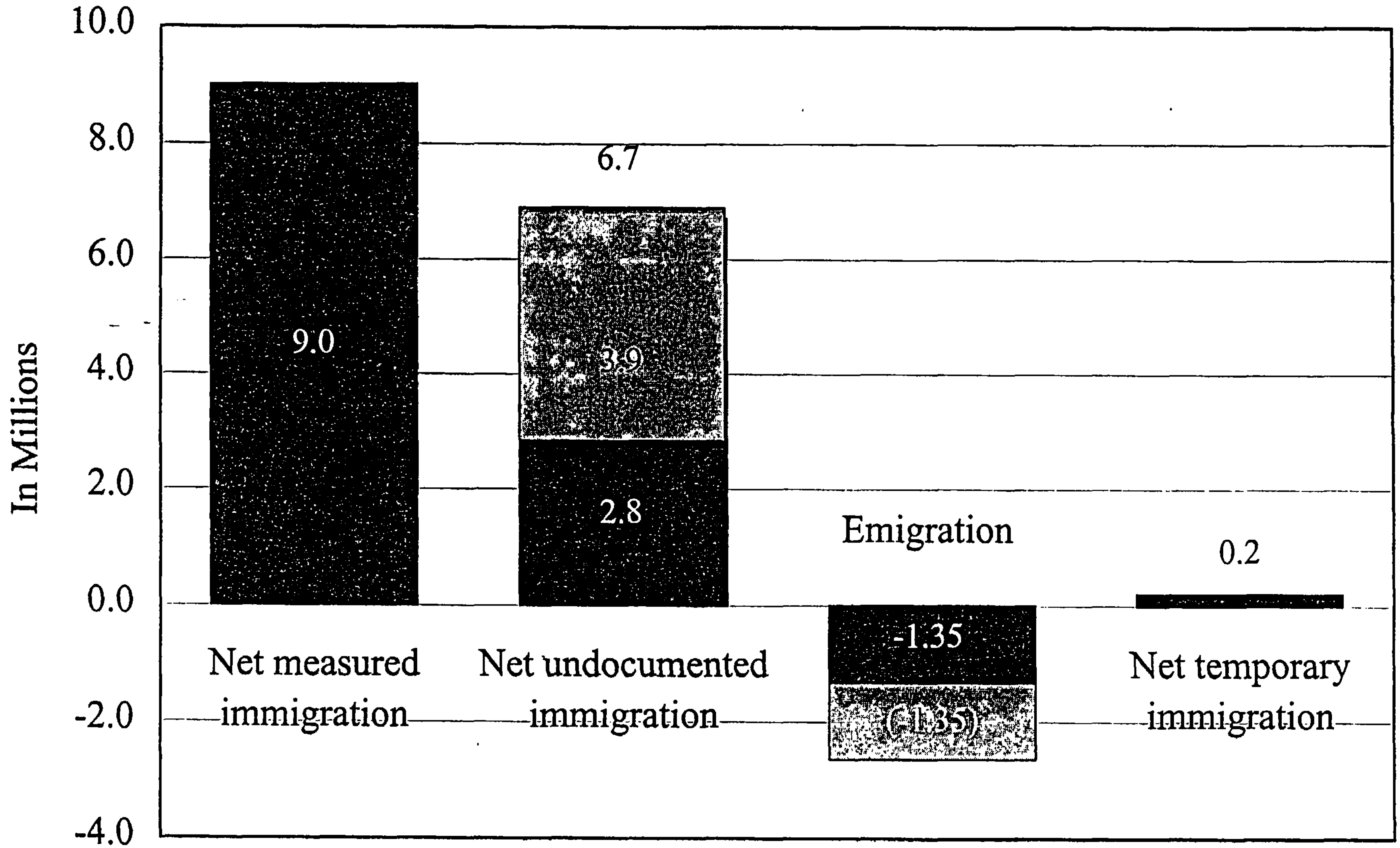
Comparison of Unexplained Change with the Results of Scenario 3 by Sex



Scenario 3 by Race and Hispanic Origin for all Age Groups



Scenario 3: Half Emigration and Remainder Undocumented



Materials attached to these minutes were draft and preliminary material to inform the ESCAP Committee. The data and analysis contained in these documents are subject to revision and are not final. These materials report the results of research and analysis undertaken by Census Bureau staff. They have undergone a more limited review than official Census Bureau publications. Research results and conclusions expressed are those of the authors and do not necessarily indicate concurrence by the Census Bureau.

Title 13-protected and/or other sensitive data, and/or detailed group quarters data that have not yet been officially released were deleted from the attached materials prior to their posting to this web page.

Coverage and Characteristics of the Population in Group Quarters
February 20, 2001

Coverage of the Group Quarters Population in Census 2000

INFORMATION DELETED

Table 1

INFORMATION DELETED

Table 2
Documentation of Assumptions for Independent Estimates of Coverage of Group Quarters in Census 2000

Correctional Facilities

"The U.S. Department of Justice, Bureau of Justice Statistics (BJS) reported that State and Federal prison authorities had under their jurisdiction 1,322,721 inmates at year end 1999: 1,284,894 were physically in their custody. Local jails held or supervised 687,973 persons awaiting trial or serving a sentence at midyear 1999." This yields a total of 1,972,867.

Source: Corrections Statistics, Internet release on the BJS web page.

Nursing Homes

The 1997 National Nursing Home Survey, conducted by National Center for Health Statistics, reported there were:

- 1.8 million beds with an occupancy rate of 88 percent, and
- 1.6 million residents in nursing homes.

Source: National Center for Health Statistics

Juvenile Institutions

The Juvenile Offenders and Victims National Report indicated that 14 to 17 year old youth are responsible for two-thirds of juvenile arrests. To reach the potential number of youth who might be in the universe in 2000 the following was estimated:

Ages	1990	1999 Estimates
14	3,243,107	3,774,164
15	3,321,609	3,893,105
16	3,304,890	3,920,044
17	3,410,062	3,930,102
Total	13,279,668	15,517,415

The number of youth counted in juvenile facilities in the 1990 census was 104,200 or .007847 of youth in these four ages. This yields and expected number of youth in juvenile facilities in 2000 of 121765 (15,517,415 * .007847).

College Dormitories

We estimated a five percent increase in the capacity of college dormitories in 2000, yielding an estimated population of 2,051,236.

Military Quarters

In 1998, there were 1,412,000 active duty military. Of those, 1,011,811 (72 percent) were on active duty in the Continental United States, Alaska, and Hawaii.

In 1990, there were 2,079,000 active duty military. This yields as estimated number in of 1,497,000 (2,019 million * .72) on active duty in the Continental United States, Alaska, and Hawaii.

In the 1990 census, 589,700 people were tabulated at military group quarters or approximately 39 percent of the active duty personnel in 1990

In 2000, the expected number of people to be tabulated at military group quarters is 394,290.
(1,011,811 active duty in 1998) * (39 percent of active duty personnel tabulated at military quarters in 1990) = 394,290

Source: Washington Headquarters Services - Directorate for Information Operations and Reports, Internet release, Defense Data Manpower Center.

Emergency and Transitional Shelters

In 1996, the estimated the number of beds available through emergency shelters, transitional shelters, and voucher distributions was 467,000. In 1988, the number of beds was 275,000.

Source: Analysis of data from the National Survey of Homeless Assistance Providers and Clients by the Urban Institute, America's Homeless II, Populations and Services, Internet release.

Group Homes

The estimate used the 1990 count of the group home population plus the addition of 50,000 small group homes with an average of five people per household.

All Other Group Quarters

Estimates for all remaining group quarters was obtained by taking the percentage of the total population in 1990 and applying that percentage to the projected population in 2000 (274,634,000).

Table 2 (cont)

Group Quarters	1990 Population	Percent of 1990 Pop	2000 Estimated POP
Mental Hospitals	128530	.0517	141927
Chronically Ill Hosp.	40980	.0165	45252
Schools or wards for drug/alcohol abuse	20129	.0081	22227
School, hosp, or wards for the mentally retarded	103713	.0417	114523
School, hosp, or wards for the physically handicapped	20654	.0083	22807
Wards for people with no usual home elsewhere	28669	.0115	31657
Religious Group Quarters	61473	.0247	67881
Agriculture Workers' Dormitories	35280	.0142	38957
Other Workers' Dorms	22920	.0092	25309
Crews of Maritime Vessels	5658	.0023	6248
Other nonhousehold living situations	97223	.0393	107909
Staff residents of institutions	18044	.0073	19925
Natural Disasters	311	.0001	343

Table 3

INFORMATION DELETED

Table 4

INFORMATION DELETED

Table 4 (cont)

2/11/01		
Hispanic Orgin by Total Group Quarters		
Hispanic Origin	**Number**	**Percent**
Total	7778633	100
NOT Hispanic	7065368	90.8
Hispanic	713265	9.2

filename - hispanicoriginbygq

2/18/01

Table 3a

Group Quarter by Race (IMPRACE variable), Age, and Sex

All Group Quarters

All persons	Total POP 1/	Total GQ	Pct GQ	Pct Total POP	0 to 17 years Total POP	GQ	Pct GQ	Pct Total	18 to 29 years Total POP	GQ	Pct. GQ	Pct Total	30 to 49 years Total POP	GQ	Pct. GQ	Pct Total	50 and over years Total POP	GQ	Pct GQ	Pct Total
White Alone	228782953	5642890	72.5	2.5	55932730	186219	57.7	0.3	36546944	2378309	71.1	6.5	69515776	956310	57.6	1.4	66457503	2122052	86.5	3.2
Black Alone	35372182	1747758	22.5	4.9	11112060	110687	34.3	1.0	8519531	725055	21.7	11.1	10844189	628139	37.9	5.8	7096402	283877	11.6	4.0
AIAN Alone	2450499	77226	1.0	3.2	818374	8829	2.7	1.1	470918	31480	0.9	6.7	728599	25284	1.5	3.5	432310	11633	0.5	2.7
Asian Alone	11365743	210479	2.7	1.8	2831336	7119	2.2	0.3	2379192	150660	4.5	6.6	3798666	26598	1.6	0.8	2378549	26102	1.1	1.2
NHOPI Alone	0	12094	0.2		0	835	0.3		0	7034	0.2		0	2841	0.2		0	1384	0.1	
Two or more Races	3450529	88186	1.1	2.6	1599312	9222	2.9	0.6	808207	84028	1.8	8.9	785790	16559	1.0	2.2	477220	8377	0.3	1.8
Total	281421906	7778633	100.0		72293812	322911	100.0		46524790	3346566	100.0		85761820	1655731	100.0		76851984	2453425	100.0	
All Group Quarters by Race, Age, and Sex																				
Males	Total POP	Total GQ	Percent	Pct Total POP	0 to 17 years Total POP	GQ	Percent	Pct Total	18 to 29 years Total POP	GQ	Percent	Pct Total	30 to 49 years Total POP	GQ	Percent	Pct Total	50 and over years Total POP	GQ	Percent	Pct Total
White Alone	112767839	2990542	66.4	2.7	28738096	123409	57.4	0.4	18770676	1411306	68.0	7.5	35096527	765501	56.6	2.2	30162540	689324	80.2	2.3
Black Alone	16837904	1282259	28.5	7.6	5647971	75889	35.3	1.3	3171590	525147	25.3	16.6	5010352	533965	39.5	10.7	3007991	147258	17.1	4.9
AIAN Alone	1217859	54233	1.2	4.5	417193	5318	2.5	1.3	241245	22238	1.1	9.2	358526	19943	1.5	5.6	200695	6734	0.8	3.4
Asian Alone	5536911	114467	2.5	2.1	1446629	4400	2.0	0.3	1190850	80676	3.9	7.2	1823548	18360	1.4	1.1	1075684	11031	1.3	1.1
NHOPI Alone	0	7983	0.2			517	0.2			4635	0.2			2212	0.2			719	0.1	
Two or more Races	1693052	52716	1.2	3.1	809306	5309	2.5	0.7	298228	30611	1.5	10.3	370121	12463	0.9	3.4	215395	4133	0.5	1.9
Total	138053565	4502200	100.0		37059197	214842	100.0		23672589	2074715	100.0		42659074	1353444	100.0		34662705	859189	100.0	
All Group Quarters by Race, Age, and Sex																				
Females	Total POP	Total GQ	Percent	Pct. Total POP	0 to 17 years Total POP	GQ	Percent	Pct Total	18 to 29 years Total POP	GQ	Percent	Pct Total	30 to 49 years Total POP	GQ	Percent	Pct total	50 and over years Total POP	GQ	Percent	Pct Total
White Alone	115995114	2652348	81.0	2.3	27194634	62810	58.1	0.2	17776268	967001	76.0	5.4	34719249	189809	62.8	0.5	36304963	1432728	89.9	3.9
Black Alone	18534278	465499	14.2	2.5	5464089	34798	32.2	0.6	5347941	189908	15.7	6.0	5633837	94174	31.2	1.7	4088411	136619	8.6	3.3
AIAN Alone	1232640	22993	0.7	1.9	401181	3511	3.2	0.9	229671	9242	0.7	4.0	370373	5341	1.8	1.4	231415	4899	0.3	2.1
Asian Alone	5848832	96012	5.9	1.6	1384707	2719	2.5	0.2	1185342	89984	5.8	6.1	1973118	8238	2.7	0.4	1302665	18071	0.9	1.2
NHOPI Alone	0	4111	0.1			318	0.3			2499	0.2			629	0.2			665	0.0	
Two or more Races	1757477	36470	1.1	2.0	790004	3913	3.6	0.5	309979	23217	1.8	7.5	395689	4096	1.4	1.0	261825	4244	0.3	1.6
Total	143368341	3276433	100.0		35234615	108069	100.0		22852201	1271851	100.0		43092248	302287	100.0		42189279	1594226	100.0	

1/ For purposes of the comparison, we allocated total US Hispanic or Latino of any race (35 million) to each race according to the following 94 % to White alone 3 % to Black alone 1 % to American Indian and Alaska Native alone, 1 5 % to Asian alone, and 5 % to Two or more races.

filename - gqtotrace4ages

Table 5b

INFORMATION DELETED

Table 6

INFORMATION DELETED

Materials attached to these minutes were draft and preliminary material to inform the ESCAP Committee. The data and analysis contained in these documents are subject to revision and are not final. These materials report the results of research and analysis undertaken by Census Bureau staff. They have undergone a more limited review than official Census Bureau publications. Research results and conclusions expressed are those of the authors and do not necessarily indicate concurrence by the Census Bureau.

Example Illustrating the Effect of Reinstated Cases on DSE and CCF

Expected Estimates		DSE with Late Adds	Late Adds	DSE without Late Adds
Census Count	C	1,000,000		1,000,000
less				
Late Census Additions	LA	0	10,000	10,000
Whole person Imputations	II	30,000	-1,000	29,000
equals				
Data Defined	**DD**	**970,000**		**961,000**
E-sample (Expected Value)	NE	970,000		961,000
less				
Erroneous & Incomplete Enumerations	EE	48,500	-4,000	44,500
equals				
Correct Enumerations	**CE**	**921,500**		**916,500**
P-Sample (ACE) Total	NP	950,000		950,000
less				
Non-matches	NM	95,000	4,639	99,639
equals				
Matches	M	855,000	-4,639	850,361
Census Coverage Ratio	**M/NP**	**0.9000**		**0.8951**
ACE Coverage Ratio	*M/CE*	*0.9278*		*0.9278*
Estimated True Population DSE	CE*NP/M	**1,023,889**		**1,023,889**
Census Total	C	1,000,000		1,000,000
Coverage Correction Factor	CCF	1.0239		1.0239
Net Undercount	DSE - C	23,889		23,889
Net Percent Undercount	1-C/DSE	2.33		2.33

Materials attached to these minutes were draft and preliminary material to inform the ESCAP Committee. The data and analysis contained in these documents are subject to revision and are not final. These materials report the results of research and analysis undertaken by Census Bureau staff. They have undergone a more limited review than official Census Bureau publications. Research results and conclusions expressed are those of the authors and do not necessarily indicate concurrence by the Census Bureau.

Table 1: Census 2000 A.C.E. 64 Post-Stratum Groups - Percent Late Adds in Hhld Pop

Race/Hispanic Origin Domain Number*	Tenure	MSA/TEA	High Return Rate NE	High Return Rate MW	High Return Rate S	High Return Rate W	Low Return Rate NE	Low Return Rate MW	Low Return Rate S	Low Return Rate W
Domain 7 (Non-Hispanic White or "Some other race")	Owner	Large MSA MO/MB	0.52	0.23	0.21	0.28	2.05	1.38	0.33	0.51
		Medium MSA MO/MB	0.37	0.23	0.26	0.28	1.02	0.55	0.36	0.37
		Small MSA & Non-MSA MO/MB	0.41	0.27	0.29	0.32	0.75	0.43	0.42	0.48
		All Other TEAs	1.32	1.29	1.57	1.41	1.88	1.90	2.26	2.24
	Non-Owner	Large MSA MO/MB	0.80				1.35			
		Medium MSA MO/MB	0.71				0.76			
		Small MSA & Non-MSA MO/MB	0.76				0.80			
		All Other TEAs	2.53				3.02			
Domain 4 (Non-Hispanic Black)	Owner	Large MSA MO/MB	0.42				1.82			
		Medium MSA MO/MB								
		Small MSA & Non-MSA MO/MB	1.24				2.24			
		All Other TEAs								
	Non-Owner	Large MSA MO/MB	0.70				1.35			
		Medium MSA MO/MB								
		Small MSA & Non-MSA MO/MB	1.11				1.56			
		All Other TEAs								
Domain 3 (Hispanic)	Owner	Large MSA MO/MB	0.50				1.58			
		Medium MSA MO/MB								
		Small MSA & Non-MSA MO/MB	1.02				2.20			
		All Other TEAs								
	Non-Owner	Large MSA MO/MB	0.69				1.35			
		Medium MSA MO/MB								
		Small MSA & Non-MSA MO/MB	1.39				2.46			
		All Other TEAs								
Domain 5 (Native Hawaiian or Pacific Islander)	Owner		0.87							
	Non-Owner		0.92							
Domain 6 (Non-Hispanic Asian)	Owner		0.69							
	Non-Owner		0.77							
American Indian or Alaska Native — **Domain 1** (On Reservation)	Owner		0.97							
	Non-Owner		0.94							
American Indian or Alaska Native — **Domain 2** (Off Reservation)	Owner		1.20							
	Non-Owner		1.16							

- For Census 2000, persons can self-identify with more than one race group. For post-stratification purposes, persons are included in a single Race/Hispanic Origin Domain. This classification does not change a person's actual response Further, all official tabulations are based on actual responses to the census.

Table 2: Census 2000 A.C.E. 64 Post-Stratum Groups - Late Adds (in Thousands)

<table>
<tr><th colspan="2" rowspan="2">Race/Hispanic Origin Domain Number*</th><th rowspan="2">Tenure</th><th rowspan="2">MSA/TEA</th><th colspan="4">High Return Rate</th><th colspan="4">Low Return Rate</th></tr>
<tr><th>NE</th><th>MW</th><th>S</th><th>W</th><th>NE</th><th>MW</th><th>S</th><th>W</th></tr>
<tr><td colspan="2" rowspan="8">Domain 7
(Non-Hispanic White or "Some other race")</td><td rowspan="4">Owner</td><td>Large MSA MO/MB</td><td>59.7</td><td>15.6</td><td>11.2</td><td>16.4</td><td>62.4</td><td>9.8</td><td>5.0</td><td>4.4</td></tr>
<tr><td>Medium MSA MO/MB</td><td>20.6</td><td>26.9</td><td>31.4</td><td>22.9</td><td>4.7</td><td>6.5</td><td>13.0</td><td>8.3</td></tr>
<tr><td>Small MSA & Non-MSA MO/MB</td><td>11.7</td><td>30.6</td><td>22.6</td><td>11.5</td><td>2.9</td><td>4.0</td><td>14.4</td><td>7.0</td></tr>
<tr><td>All Other TEAs</td><td>54.5</td><td>89.3</td><td>78.4</td><td>25.9</td><td>26.6</td><td>20.6</td><td>269.7</td><td>52.2</td></tr>
<tr><td rowspan="4">Non-Owner</td><td>Large MSA MO/MB</td><td colspan="4">68.5</td><td colspan="4">51.3</td></tr>
<tr><td>Medium MSA MO/MB</td><td colspan="4">81.7</td><td colspan="4">23.6</td></tr>
<tr><td>Small MSA & Non-MSA MO/MB</td><td colspan="4">71.9</td><td colspan="4">17.2</td></tr>
<tr><td>All Other TEAs</td><td colspan="4">126.5</td><td colspan="4">75.2</td></tr>
<tr><td colspan="2" rowspan="8">Domain 4
(Non-Hispanic Black)</td><td rowspan="4">Owner</td><td>Large MSA MO/MB</td><td colspan="4" rowspan="2">35.3</td><td colspan="4" rowspan="2">52.0</td></tr>
<tr><td>Medium MSA MO/MB</td></tr>
<tr><td>Small MSA & Non-MSA MO/MB</td><td colspan="4" rowspan="2">48.4</td><td colspan="4" rowspan="2">29.1</td></tr>
<tr><td>All Other TEAs</td></tr>
<tr><td rowspan="4">Non-Owner</td><td>Large MSA MO/MB</td><td colspan="4" rowspan="2">64.8</td><td colspan="4" rowspan="2">50.2</td></tr>
<tr><td>Medium MSA MO/MB</td></tr>
<tr><td>Small MSA & Non-MSA MO/MB</td><td colspan="4" rowspan="2">38.7</td><td colspan="4" rowspan="2">8.2</td></tr>
<tr><td>All Other TEAs</td></tr>
<tr><td colspan="2" rowspan="8">Domain 3
(Hispanic)</td><td rowspan="4">Owner</td><td>Large MSA MO/MB</td><td colspan="4" rowspan="2">48.3</td><td colspan="4" rowspan="2">39.2</td></tr>
<tr><td>Medium MSA MO/MB</td></tr>
<tr><td>Small MSA & Non-MSA MO/MB</td><td colspan="4" rowspan="2">29.6</td><td colspan="4" rowspan="2">36.6</td></tr>
<tr><td>All Other TEAs</td></tr>
<tr><td rowspan="4">Non-Owner</td><td>Large MSA MO/MB</td><td colspan="4" rowspan="2">74.0</td><td colspan="4" rowspan="2">51.0</td></tr>
<tr><td>Medium MSA MO/MB</td></tr>
<tr><td>Small MSA & Non-MSA MO/MB</td><td colspan="4" rowspan="2">35.6</td><td colspan="4" rowspan="2">16.6</td></tr>
<tr><td>All Other TEAs</td></tr>
<tr><td colspan="2" rowspan="2">Domain 5
(Native Hawaiian or Pacific Islander)</td><td colspan="2">Owner</td><td colspan="8">2.7</td></tr>
<tr><td colspan="2">Non-Owner</td><td colspan="8">2.6</td></tr>
<tr><td colspan="2" rowspan="2">Domain 6
(Non-Hispanic Asian)</td><td colspan="2">Owner</td><td colspan="8">41.8</td></tr>
<tr><td colspan="2">Non-Owner</td><td colspan="8">30.2</td></tr>
<tr><td rowspan="4">American Indian or Alaska Native</td><td rowspan="2">Domain 1
(On Reservation)</td><td colspan="2">Owner</td><td colspan="8">3.6</td></tr>
<tr><td colspan="2">Non-Owner</td><td colspan="8">1.6</td></tr>
<tr><td rowspan="2">Domain 2
(Off Reservation)</td><td colspan="2">Owner</td><td colspan="8">11.1</td></tr>
<tr><td colspan="2">Non-Owner</td><td colspan="8">7.5</td></tr>
</table>

• For Census 2000, persons can self-identify with more than one race group. For post-stratification purposes, persons are included in a single Race/Hispanic Origin Domain. This classification does not change a person's actual response. Further, all official tabulations are based on actual responses to the census.

Table 3: Census 2000 A.C.E. 64 Post-Stratum Groups - Percent Net Undercount

Race/Hispanic Origin Domain Number*	Tenure	MSA/TEA	High Return Rate				Low Return Rate			
			NE	MW	S	W	NE	MW	S	W
Domain 7 (Non-Hispanic White or "Some other race")	Owner	Large MSA MO/MB	0.81	0.01	0.36	-0.38	-3.62	-2.61	2.19	1.14
		Medium MSA MO/MB	0.30	-0.12	0.46	-0.28	-4.39	-0.33	0.66	1.81
		Small MSA & Non-MSA MO/MB	-0.25	0.14	0.44	0.30	2.29	2.61	2.09	2.71
		All Other TEAs	1.84	-1.11	1.34	0.85	0.56	-0.16	0.15	1.59
	Non-Owner	Large MSA MO/MB	1.82				1.02			
		Medium MSA MO/MB	0.61				2.83			
		Small MSA & Non-MSA MO/MB	2.45				3.61			
		All Other TEAs	1.64				4.08			
Domain 4 (Non-Hispanic Black)	Owner	Large MSA MO/MB	1.63				-1.31			
		Medium MSA MO/MB								
		Small MSA & Non-MSA MO/MB	0.07				0.46			
		All Other TEAs								
	Non-Owner	Large MSA MO/MB	4.18				3.42			
		Medium MSA MO/MB								
		Small MSA & Non-MSA MO/MB	2.64				0.12			
		All Other TEAs								
Domain 3 (Hispanic)	Owner	Large MSA MO/MB	1.46				0.04			
		Medium MSA MO/MB								
		Small MSA & Non-MSA MO/MB	1.66				1.08			
		All Other TEAs								
	Non-Owner	Large MSA MO/MB	3.52				4.98			
		Medium MSA MO/MB								
		Small MSA & Non-MSA MO/MB	4.88				10.74			
		All Other TEAs								
Domain 5 (Native Hawaiian or Pacific Islander)	Owner		2.71							
	Non-Owner		6.58							
Domain 6 (Non-Hispanic Asian)	Owner		0.55							
	Non-Owner		1.58							
American Indian or Alaska Native – **Domain 1** (On Reservation)	Owner		5.04							
	Non-Owner		4.10							
American Indian or Alaska Native – **Domain 2** (Off Reservation)	Owner		1.60							
	Non-Owner		5.57							

- For Census 2000, persons can self-identify with more than one race group. For post-stratification purposes, persons are included in a single Race/Hispanic Origin Domain. This classification does not change a person's actual response. Further, all official tabulations are based on actual responses to the census

Table 4: Census 2000 A.C.E. 64 Post-Stratum Groups -Net Undercount (Thousands)

Race/Hispanic Origin Domain Number*	Tenure	MSA/TEA	High Return Rate				Low Return Rate			
			NE	MW	S	W	NE	MW	S	W
Domain 7 (Non-Hispanic White or "Some other race")	Owner	Large MSA MO/MB	93.3	0.6	19.5	-22.2	-106.2	-17.9	34.1	10.0
		Medium MSA MO/MB	16.7	-14.5	55.3	-22.9	-19.3	-3.8	23.7	42.1
		Small MSA & Non-MSA MO/MB	-7.2	16.4	34.8	10.8	9.1	24.9	73.0	40.8
		All Other TEAs	77.3	-76.0	68.3	15.7	8.0	-1.7	18.0	37.6
	Non-Owner	Large MSA MO/MB	158.0				39.1			
		Medium MSA MO/MB	70.3				90.5			
		Small MSA & Non-MSA MO/MB	236.8				80.4			
		All Other TEAs	83.1				106.1			
Domain 4 (Non-Hispanic Black)	Owner	Large MSA MO/MB	140.9				-36.8			
		Medium MSA MO/MB								
		Small MSA & Non-MSA MO/MB	2.9				6.0			
		All Other TEAs								
	Non-Owner	Large MSA MO/MB	401.4				131.4			
		Medium MSA MO/MB								
		Small MSA & Non-MSA MO/MB	94.4				0.6			
		All Other TEAs								
Domain 3 (Hispanic)	Owner	Large MSA MO/MB	144.2				0.9			
		Medium MSA MO/MB								
		Small MSA & Non-MSA MO/MB	49.0				18.2			
		All Other TEAs								
	Non-Owner	Large MSA MO/MB	391.6				198.0			
		Medium MSA MO/MB								
		Small MSA & Non-MSA MO/MB	130.9				81.2			
		All Other TEAs								
Domain 5 (Native Hawaiian or Pacific Islander)	Owner		8.5							
	Non-Owner		20.0							
Domain 6 (Non-Hispanic Asian)	Owner		33.2							
	Non-Owner		63.2							
American Indian or Alaska Native – **Domain 1** (On Reservation)	Owner		19.5							
	Non-Owner		7.4							
American Indian or Alaska Native – **Domain 2** (Off Reservation)	Owner		15.0							
	Non-Owner		38.0							

* For Census 2000, persons can self-identify with more than one race group. For post-stratification purposes, persons are included in a single Race/Hispanic Origin Domain. This classification does not change a person's actual response. Further, all official tabulations are based on actual responses to the census

Table 5: Census 2000 A.C.E. 64 Post-Stratum Groups -Census Counts (Millions)

Race/Hispanic Origin Domain Number*	Tenure	MSA/TEA	High Return Rate				Low Return Rate			
			NE	MW	S	W	NE	MW	S	W
Domain 7 (Non-Hispanic White or "Some other race")	Owner	Large MSA MO/MB	11.5	6 8	5 3	5.8	3 0	0.7	1 5	0 9
		Medium MSA MO/MB	5.6	11 7	11.9	8.2	0 5	1 2	3 6	2 3
		Small MSA & Non-MSA MO/MB	2.9	11 4	7 9	3 6	0 4	0.9	3 4	1 5
		All Other TEAs	4 1	6.9	5 0	1.8	1 4	1.0	11 9	2 3
	Non-Owner	Large MSA MO/MB	8 5				3.8			
		Medium MSA MO/MB	11.5				3.1			
		Small MSA & Non-MSA MO/MB	9.4				2.1			
		All Other TEAs	5 0				2.5			
Domain 4 (Non-Hispanic Black)	Owner	Large MSA MO/MB; Medium MSA MO/MB	8.5				2.8			
		Small MSA & Non-MSA MO/MB; All Other TEAs	3 9				1.3			
	Non-Owner	Large MSA MO/MB; Medium MSA MO/MB	9.2				3.7			
		Small MSA & Non-MSA MO/MB; All Other TEAs	3.5				0 5			
Domain 3 (Hispanic)	Owner	Large MSA MO/MB; Medium MSA MO/MB	9 7				2.5			
		Small MSA & Non-MSA MO/MB; All Other TEAs	2 9				1 7			
	Non-Owner	Large MSA MO/MB; Medium MSA MO/MB	10 7				3.8			
		Small MSA & Non-MSA MO/MB; All Other TEAs	2.6				0.7			
Domain 5 (Native Hawaiian or Pacific Islander)	Owner		0.3							
	Non-Owner		0.3							
Domain 6 (Non-Hispanic Asian)	Owner		6.0							
	Non-Owner		3.9							
American Indian or Alaska Native — **Domain 1** (On Reservation)	Owner		0 4							
	Non-Owner		0.2							
American Indian or Alaska Native — **Domain 2** (Off Reservation)	Owner		0.9							
	Non-Owner		0.6							

* For Census 2000, persons can self-identify with more than one race group. For post-stratification purposes, persons are included in a single Race/Hispanic Origin Domain. This classification does not change a person's actual response. Further, all official tabulations are based on actual responses to the census

ESCAP MEETING NO. 42 - 02/20/01

MINUTES

Minutes of the Executive Steering Committee on
Accuracy and Coverage Evaluation (A.C.E.) Policy (ESCAP) Meeting # 42

February 20, 2001

Prepared by: Sarah Brady

The forty-second meeting of the Executive Steering Committee on Accuracy and Coverage Evaluation Policy was held on February 20, 2001 at 10:30 and at 1:30. The agenda for the 10:30 meeting was to discuss scenarios for Demographic Analysis and to present Group Quarters data. The agenda for the afternoon was to present data for late census adds and to discuss expected loss.

Committee Attendees:

Nancy Potok
Paula Schneider
Cynthia Clark
Nancy Gordon
John Thompson
Jay Waite
Bob Fay
Howard Hogan
Ruth Ann Killion
John Long
Carol Van Horn

Deputy Director/Acting Director:
William Barron

Other Attendees:

Marvin Raines
Bill Bell (PM only)
Kathleen Styles (AM only)
Maria Urrutia
Sarah Brady

I. Scenarios for Demographic Analysis

John Long presented results from Scenario 3 of the demographic analysis (DA) alternative scenarios to explain the difference between adjusted 1990 (PES) and adjusted 2000 (A.C.E.). Scenario 3 allows for a 50% reduction in emigration with the remaining unexplained difference between the PES and A.C.E. in net undocumented immigration. The Committee noted that although scenario 3 accounted for some of the unexplained difference between the A.C.E. and DA, it did not explain all of the difference. The attached handout discusses all of the scenarios proposed to explain the difference between A.C.E. and DA. For item 2 on the handout the priority is given to examining the 1990 unadjusted census to unadjusted Census 2000.

II. Group Quarters

John Long also presented data for the group quarters (GQ) population. The data compared the GQ population as measured by the Census 2000 to independent estimates. The Committee discussed that while some differences are seen between types of GQs, overall, the results were what we expected. The data examined indicated that the difference between the A.C.E. and DA is not due to an overcount in the group quarters population.

Jay Waite requested that data be tabulated to examine if there was an unusual increase in college age people living at home as compared to 1990. This was requested to see if there was large scale duplication between group quarters (specifically colleges) and housing units which could account for some of the difference between the A.C.E. and DA results.

III. Distribution of Late Census Data

Howard Hogan presented data for late census adds. Howard first discussed the underlying assumption of how late census adds are treated in the A.C.E. The key assumption that ensures that late adds do not affect the A.C.E. estimates of coverage error is that the A.C.E. probability of capture for correct enumerations in the late census data universe is the same as for correct enumerations not in late census data universe. Howard then presented an example illustrating this assumption.

Howard distributed the attached handout of the distribution of late adds by post-stratum groups. If the late adds tended to cluster in the post-strata, it indicated clustering of net coverage which would be an issue for our synthetic assumption. John Thompson indicated that he would like to see the distribution of late adds by region for the post-strata groups.

The late census adds were due to the unduplication operation implemented for Census 2000. The committee noted that it was fortunate we carried out the unduplication operation because it improved the quality of the census for apportionment.

On a related topic, Bob Fay indicated that he had potentially found a balancing issue. He is currently documenting the issue and it will be discussed further at a future meeting.

IV. Expected Loss

Howard Hogan presented information about expected loss. For the loss functions, what is known about the A.C.E. and it's biases are used to derive a "true" population. The A.C.E. and census are compared to these "truths" or targets to determine which of the two is closer to the "truth". Therefore, these targets imply expected loss for the census and A.C.E. The analytical framework for the Committee's recommendation is based on the concept of finding which expected loss is smaller, the unadjusted census or the adjusted census. This type of analysis is different from the hypothesis testing done in 1990 which assumed the unadjusted census was more accurate unless proven otherwise.

Given the uncertainty in the estimates of total error, the Committee expressed concerns regarding interpretations of the results of the loss function analysis, particularly if they were close. John Thompson noted that this is why it is important to analyze the sensitivity of the loss functions to specific parameters to see what makes a strong case for adjusting or not adjusting. The Committee also expressed a desire to study the distributions for the simulations used in calculating the expected loss to determine the closeness of the expected losses between the unadjusted census and adjusted census.

V. Next Meeting

The agenda for the next meeting, scheduled for February 21, 2001 is to discuss results from additional DA scenarios.

ESCAP MEETING NO. 43 - 02/21/01

AGENDA

There was no agenda developed or used for the February 21, 2001 meeting.

ESCAP MEETING NO. 43 - 02/21/01

HANDOUTS

Materials attached to these minutes were draft and preliminary material to inform the ESCAP Committee. The data and analysis contained in these documents are subject to revision and are not final. These materials report the results of research and analysis undertaken by Census Bureau staff. They have undergone a more limited review than official Census Bureau publications. Research results and conclusions expressed are those of the authors and do not necessarily indicate concurrence by the Census Bureau.

Table A-1: Census 2000 A.C.E. 64 Post-Stratum Groups by region - Percent Late Adds

Race/Hispanic Origin Domain Number*	Tenure	MSA/TEA	High Return Rate				Low Return Rate			
			NE	MW	S	W	NE	MW	S	W
Domain 7 (Non-Hispanic White or "Some other race")	Owner	Large MSA MO/MB	0.5	0.2	0.2	0 3	2.1	1.4	0.3	0.5
		Medium MSA MO/MB	0.4	0 2	0 3	0 3	1.0	0.6	0 4	0.4
		Small MSA & Non-MSA MO/MB	0.4	0 3	0.3	0.3	0.8	0.4	0.4	0 5
		All Other TEAs	1.3	1.3	1.6	1.4	1.9	1.9	2.3	2 2
	Non-Owner	Large MSA MO/MB	1 2	0.8	0 5	0 5	1.7	1.6	0.7	0.6
		Medium MSA MO/MB	1.3	0.7	0.6	0.6	1.8	0.9	0.6	0.6
		Small MSA & Non-MSA MO/MB	1.3	0.7	0.7	0.7	1.5	0.7	0.7	0.7
		All Other TEAs	3.2	2.2	2.4	2.7	3.8	2.5	2.8	3.5
Domain 4 (Non-Hispanic Black)	Owner	Large MSA MO/MB	0 8	0.6	0.3	0.3	3.2	1.7	0.4	0.3
		Medium MSA MO/MB								
		Small MSA & Non-MSA MO/MB	1.0	0.4	1.3	0.7	0.8	0.7	2.3	1.4
		All Other TEAs								
	Non-Owner	Large MSA MO/MB	0.9	1.0	0.6	0.5	1.6	1.7	0.8	0.6
		Medium MSA MO/MB								
		Small MSA & Non-MSA MO/MB	1.5	0.6	1.2	0.7	1.9	0.6	1.7	1.2
		All Other TEAs								
Domain 3 (Hispanic)	Owner	Large MSA MO/MB	1.2	0.5	0.4	0.4	3.3	2.7	0.6	0.5
		Medium MSA MO/MB								
		Small MSA & Non-MSA MO/MB	1.1	0.6	1.2	0.9	1.2	1.1	2.2	2.4
		All Other TEAs								
	Non-Owner	Large MSA MO/MB	1.2	1.0	0.6	0.5	1.9	2.2	0.5	0.6
		Medium MSA MO/MB								
		Small MSA & Non-MSA MO/MB	2.1	1.0	1.4	1.4	1.7	0.9	2.1	3.3
		All Other TEAs								
Domain 5 (Native Hawaiian or Pacific Islander)	Owner		1.6		0.5		0.4		0.9	
	Non-Owner		1.2		0.6		0.6		1 0	
Domain 6 (Non-Hispanic Asian)	Owner		1.8		0.5		0.4		0.5	
	Non-Owner		1.3		0.8		0.5		0.6	
American Indian or Alaska Native — **Domain 1** (On Reservation)	Owner		1.6		0.8		1.5		0.9	
	Non-Owner		0.8		0.7		1.4		1.1	
American Indian or Alaska Native — **Domain 2** (Off Reservation)	Owner		1.4		0.7		1.5		0.9	
	Non-Owner		1.7		1.0		1.3		1.0	

* For Census 2000, persons can self-identify with more than one race group For post-stratification purposes, persons are included in a single Race/Hispanic Origin Domain. This classification does not change a person's actual response. Further, all official tabulations are based on actual responses to the census.

Table A-2: Census 2000 A.C.E. 64 Post-Stratum Groups by region - Percent iis

Race/Hispanic Origin Domain Number*	Tenure	MSA/TEA	High Return Rate NE	High Return Rate MW	High Return Rate S	High Return Rate W	Low Return Rate NE	Low Return Rate MW	Low Return Rate S	Low Return Rate W
Domain 7 (Non-Hispanic White or "Some other race")	Owner	Large MSA MO/MB	0.9	0.9	1.0	1.1	2.6	2.6	1.8	1.8
		Medium MSA MO/MB	0.8	0.8	1.1	1.2	1.8	1.6	1.6	1.9
		Small MSA & Non-MSA MO/MB	0.8	0.9	0.9	1.1	1.2	1.8	1.5	1.9
		All Other TEAs	1.6	1.1	1.1	2.8	1.6	1.5	1.5	2.5
	Non-Owner	Large MSA MO/MB	2.1	2.0	2.7	1.9	4.4	5.4	3.9	3.1
		Medium MSA MO/MB	1.6	1.6	2.4	2.3	3.5	3.0	3.4	3.2
		Small MSA & Non-MSA MO/MB	1.4	1.6	2.0	1.9	2.2	3.3	2.8	2.8
		All Other TEAs	2.1	1.6	1.8	3.7	3.0	2.1	2.5	3.9
Domain 4 (Non-Hispanic Black)	Owner	Large MSA MO/MB Medium MSA MO/MB	2.7	2.4	2.4	2.5	5.3	4.9	3.3	3.4
		Small MSA & Non-MSA MO/MB All Other TEAs	2.6	2.3	2.2	2.9	2.8	3.4	3.3	5.2
	Non-Owner	Large MSA MO/MB Medium MSA MO/MB	4.4	3.3	3.5	3.1	6.6	6.0	4.4	4.5
		Small MSA & Non-MSA MO/MB All Other TEAs	2.5	3.0	2.9	3.2	3.7	5.9	3.9	5.6
Domain 3 (Hispanic)	Owner	Large MSA MO/MB Medium MSA MO/MB	2.5	2.5	2.6	4.0	5.9	5.3	4.2	5.0
		Small MSA & Non-MSA MO/MB All Other TEAs	2.4	2.3	4.0	3.9	3.8	3.5	4.1	4.5
	Non-Owner	Large MSA MO/MB Medium MSA MO/MB	3.6	3.2	3.2	3.8	5.4	5.7	4.0	4.3
		Small MSA & Non-MSA MO/MB All Other TEAs	2.5	2.8	3.7	4.1	3.9	4.0	4.9	5.4
Domain 5 (Native Hawaiian or Pacific Islander)	Owner		3.3		3.7		2.8		3.7	
	Non-Owner		5.1		4.3		4.3		3.6	
Domain 6 (Non-Hispanic Asian)	Owner		2.8		2.3		2.2		2.4	
	Non-Owner		3.8		3.3		3.3		3.1	
American Indian or Alaska Native – **Domain 1** (On Reservation)	Owner		8.8		4.9		4.2		5.0	
	Non-Owner		6.7		4.7		3.7		4.6	
American Indian or Alaska Native – **Domain 2** (Off Reservation)	Owner		2.7		2.1		1.9		3.0	
	Non-Owner		3.6		2.7		2.7		3.2	

* For Census 2000, persons can self-identify with more than one race group. For post-stratification purposes, persons are included in a single Race/Hispanic Origin Domain. This classification does not change a person's actual response Further, all official tabulations are based on actual responses to the census.

Implied Foreign-Born as a Percent of the Total Population by Alternative Scenarios

		Category			
		Totals			Asian and Pacific Islander
		US Total	Hispanic	Non-Hispanic	
2000 CPS	Total	274,087,000	32,804,000	241,283,000	10,925,000
	Foreign Born	28,379,000	12,841,000	15,538,000	6,706,000
	% of Pop	10.35	39.14	6.44	61.38
Base	Total	279,467,263	33,790,648	245,676,615	10,721,484
	Foreign Born	28,586,991	12,342,855	16,244,137	6,651,452
	% of Pop	10.23	36.53	6.61	62.04
Scenario 1	Total	284,721,906	37,590,046	247,131,860	11,084,498
	Foreign Born	33,841,635	16,142,253	17,699,382	7,014,466
	% of Pop	11.89	42.94	7.16	63.28
Scenario 2	Total	284,721,906	36,246,773	248,591,792	11,501,825
	Foreign Born	33,361,602	14,766,935	18,594,666	7,424,428
	% of Pop	11.72	40.74	7.48	64.55
Scenario 3	Total	284,721,906	36,920,358	247,801,548	11,292,810
	Foreign Born	33,601,618	15,454,594	18,147,024	7,219,447
	% of Pop	11.80	41.86	7.32	63.93
Scenario A	Total	281,421,906	35,203,963	246,217,943	10,856,519
	Foreign Born	30,541,635	13,756,170	16,785,465	6,786,487
	% of Pop	10.85	39.08	6.82	62.51
Scenario B	Total	281,421,906	34,869,119	246,552,787	10,960,675
	Foreign Born	30,421,626	13,412,341	17,009,286	6,888,978
	% of Pop	10.81	38.46	6.90	62.85
Scenario C	Total	281,421,906	36,950,125	244,471,781	10,967,985
	Foreign Born	33,797,744	16,252,529	17,545,216	6,823,304
	% of Pop	12.01	43.99	7.18	62.21
Scenario D	Total	281,421,906	36,280,437	245,141,469	11,176,297
	Foreign Born	33,787,010	15,576,646	18,210,364	7,178,789
	% of Pop	12.01	42.93	7.43	64.23

Relationship Between Census Counts, Coverage Estimates, and Demographic Change

(1) $P_{1990} + (B - D + M) = P_{2000}$

(2) $(C_{1990} + U_{1990}) + (B - D + I + N - U) = (C_{2000} + U_{2000})$

(3) $(C_{1990} + \hat{U}_{1990}) + (\hat{B} - \hat{D} + \hat{I} + \hat{N} - \hat{U}) + E = (C_{2000} + \hat{U}_{2000})$

(4) $E = (C_{2000} + \hat{U}_{2000}) - (C_{1990} + \hat{U}_{1990}) - (\hat{B} - \hat{D} + \hat{I} + \hat{N} - \hat{O})$

(5) $E = (C_{2000} - C_{1990}) + (\hat{U}_{2000} - \hat{U}_{1990}) - (\hat{B} - \hat{D} + \hat{I} + \hat{N} - \hat{O})$

Where:

P_{1990} = 1990 Population
P_{2000} = 2000 Population
B = Births 1990 – 2000
D = Deaths 1990 – 2000
M = Net Migration 1990 – 2000
C_{1990} = 1990 Census Count
C_{2000} = 2000 Census Count
U_{1990} = 1990 Undercount
U_{2000} = 2000 Undercount
I = Legal In-migration 1990 – 2000
O = Legal Out-migration 1990 – 2000
N = Net Undocumented Migration 1990 – 2000
E = Unexplained Difference

and ^ denotes estimated rather than actual value

ALTERNATE SCENARIOS

<u>I. PES Adjusted 1990 to A.C.E. Adjusted 2000</u>

(6) $(\hat{U}_{2000} - \hat{U}_{1990}) = (\hat{A}_{2000} - \hat{A}_{1990})$

where:

$\hat{A}_{2000} = \text{A.C.E.}_{2000} - C_{2000}$

$\hat{A}_{1990} = PES_{1990} - C_{1990}$

Base Scenario

$$(7)\ \hat{E} = (C_{2000} - C_{1990}) + (\hat{A}_{2000} - \hat{A}_{1990}) - (\hat{B} - \hat{D} + \hat{I} + \hat{N} - \hat{O})$$

Scenario 1 – All Unexplained Difference in Net Undocumented Immigration

$$(8)\ N_1 = \hat{N} + \hat{E}$$

Scenario 2 – No Out Migration, Remainder of Unexplained Difference in Undocumented Immigration

$$(9)\ O_2 = \varnothing, \quad N_2 = \hat{N} + (\hat{E} - \hat{O})$$

Scenario 3 – Half Out Migration, Remainder of Unexplained Difference in Undocumented Immigration

$$(10)\ O_3 = 0.5\ \hat{O}, \qquad N_3 = \hat{N} + (\hat{E} - 0.5\ \hat{O})$$

II. PES Adjusted 1990 to Unadjusted Census2000

$$(11)\ (\hat{U}_{2000} - \hat{U}_{1990}) = -\hat{A}_{1990}$$

Base Scenario

$$(12)\ \hat{E}' = (C_{2000} - C_{1990}) - \hat{A}_{1990} - (\hat{B} - \hat{D} + \hat{I} + \hat{N} - \hat{O})$$

√ *Scenario A – All Unexplained Difference (E') in Net Undocumented Immigration*

$$(13)\ N_A = \hat{N} + \hat{E}'$$

Scenario B – Reduce Out Migration by one-fourth, Remainder of Unexplained Difference (E') in Undocumented Immigration

$$(14)\quad O_B = 0.75\,\hat{O}, \qquad N_B = \hat{N} + (\hat{E}' - 0.25\,\hat{O})$$

III. Unadjusted 1990 Census to Unadjusted Census2000

$$(15)\quad (\hat{U}_{2000} - \hat{U}_{1990}) = \varnothing$$

Base Scenario

$$(16)\quad E'' = (C_{2000} - C_{1990}) - (\hat{B} - \hat{D} + \hat{I} + \hat{N} - \hat{O})$$

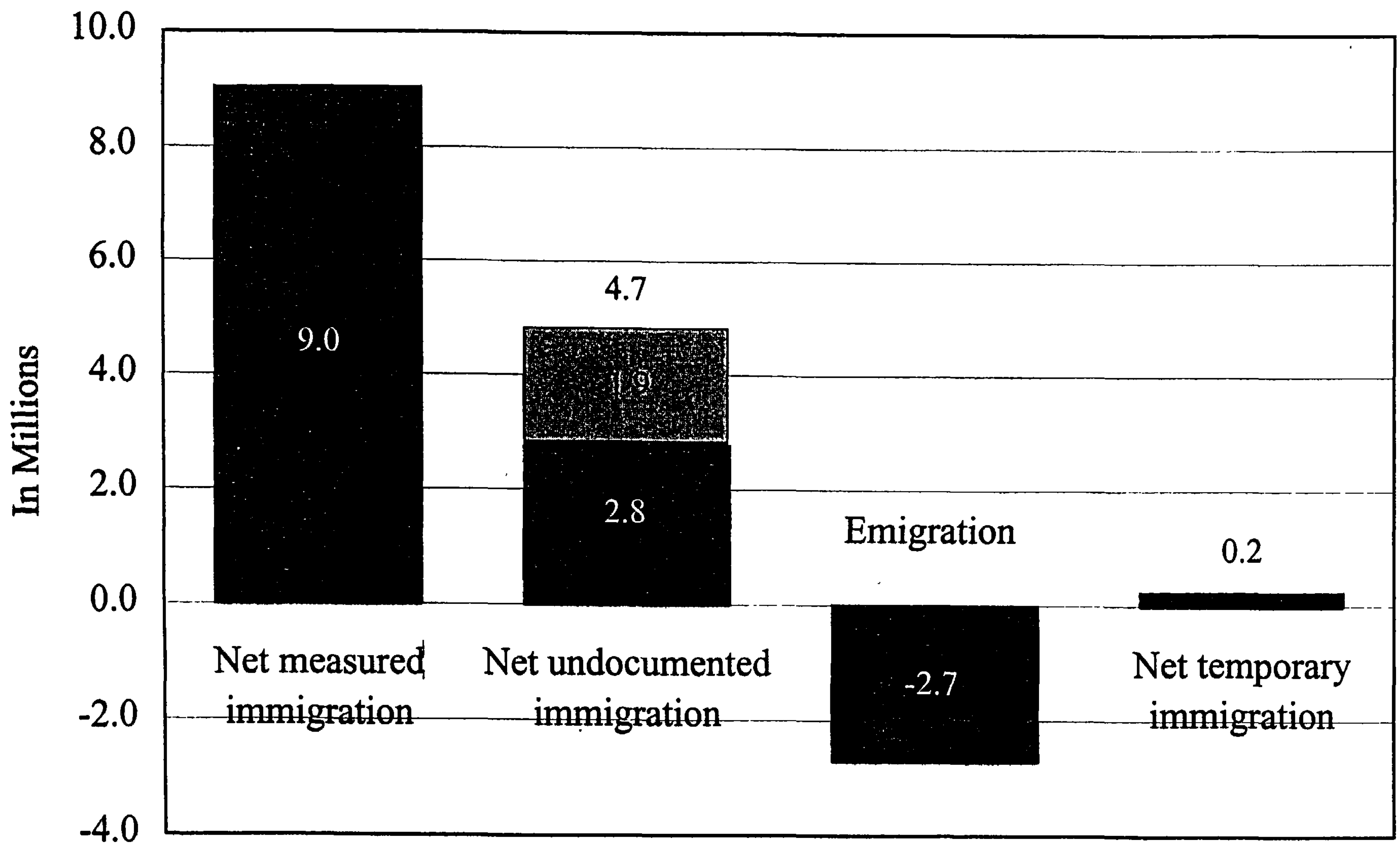
Scenario C – All Unexplained Difference (E") in Net Undocumented Immigration

$$(17)\quad N_C = \hat{N} + \hat{E}''$$

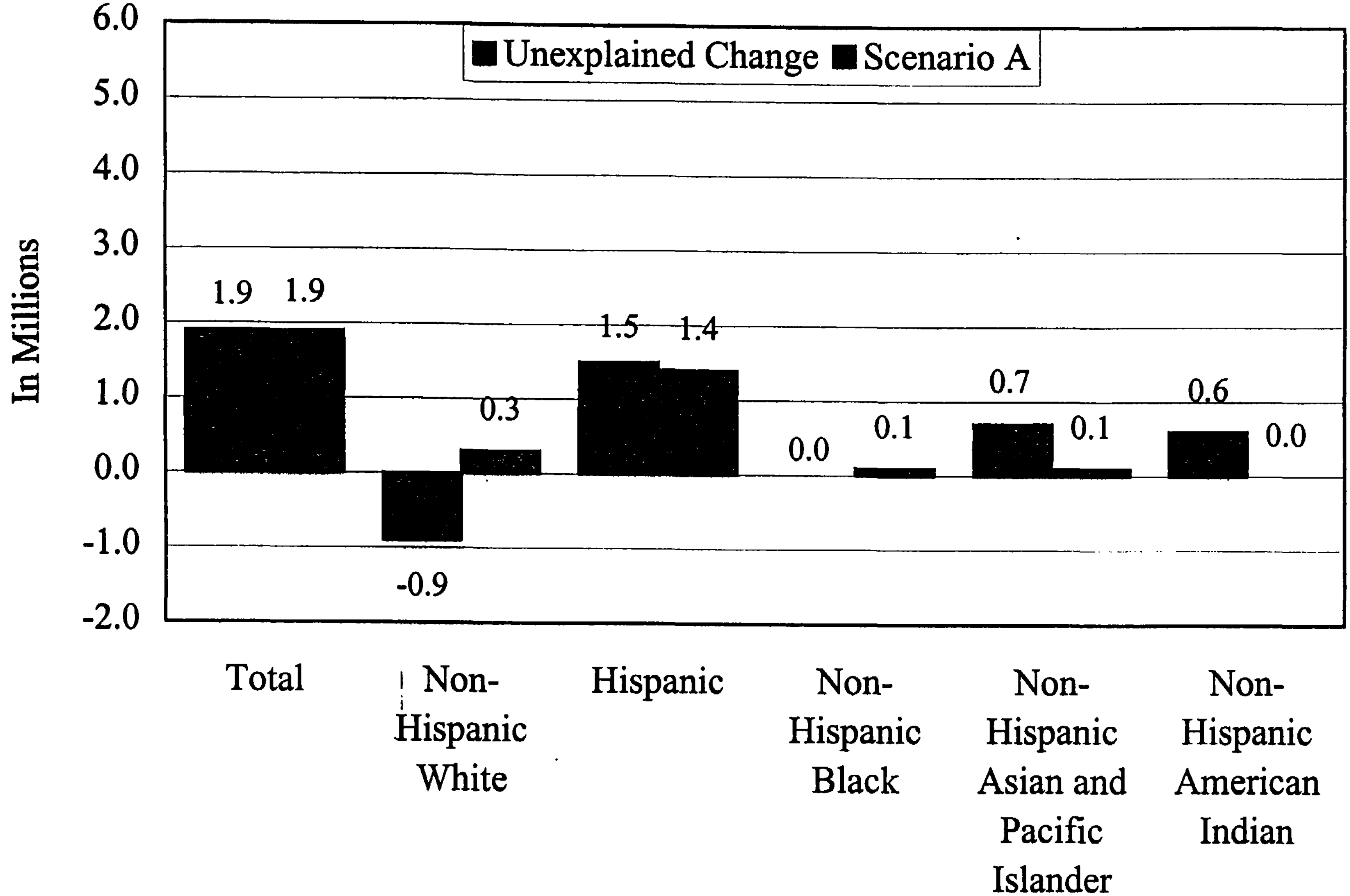
Scenario D – Half Out Migration, Remainder of Unexplained Difference (E") in Undocumented Immigration

$$(18)\quad O_D = 0.5\,\hat{O}, \qquad N_D = \hat{N} + (\hat{E}'' - 0.5\,\hat{O})$$

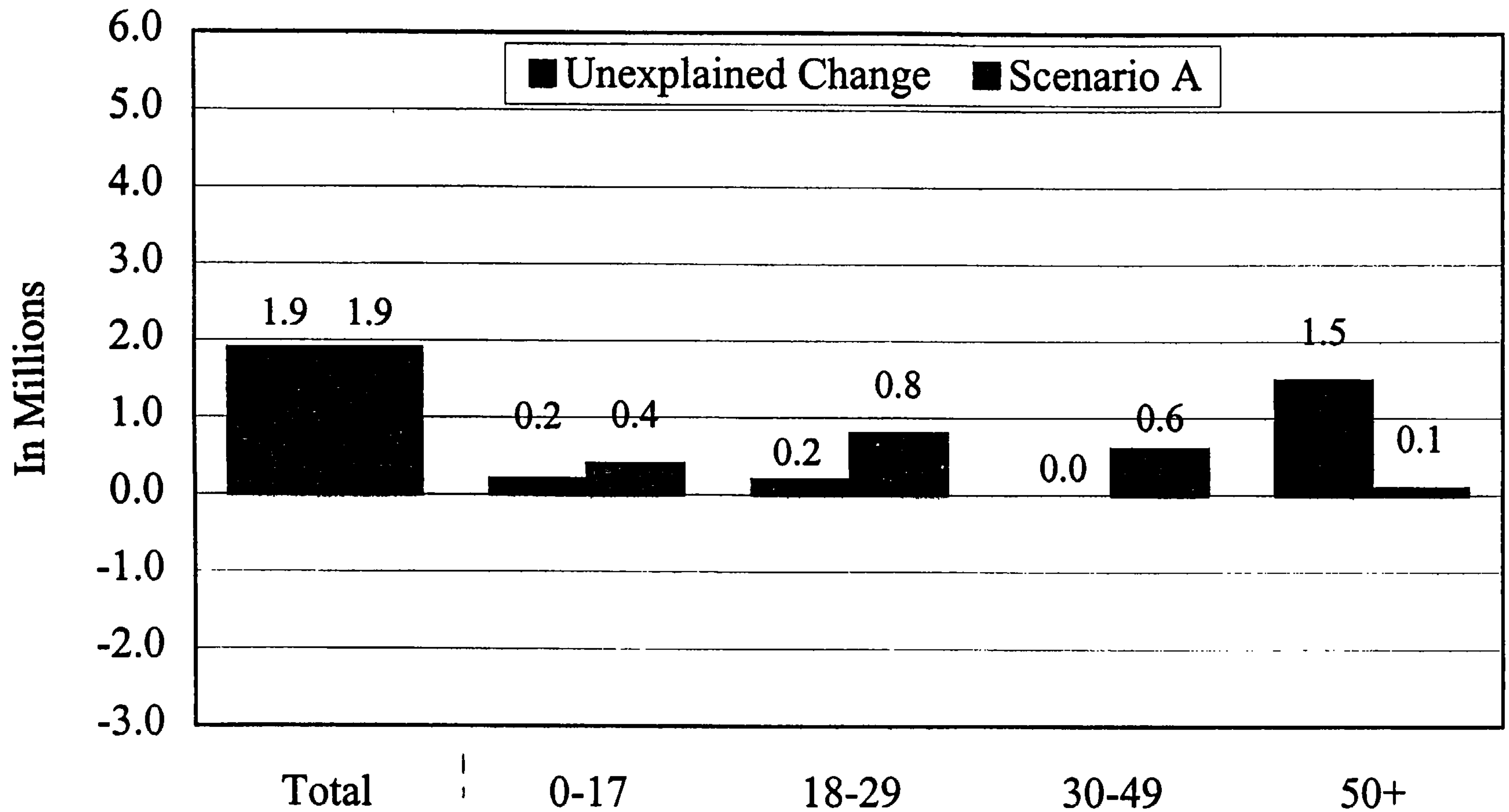
Scenario A: Increase Undocumented Migration (1990 Adj. to Census 2000)



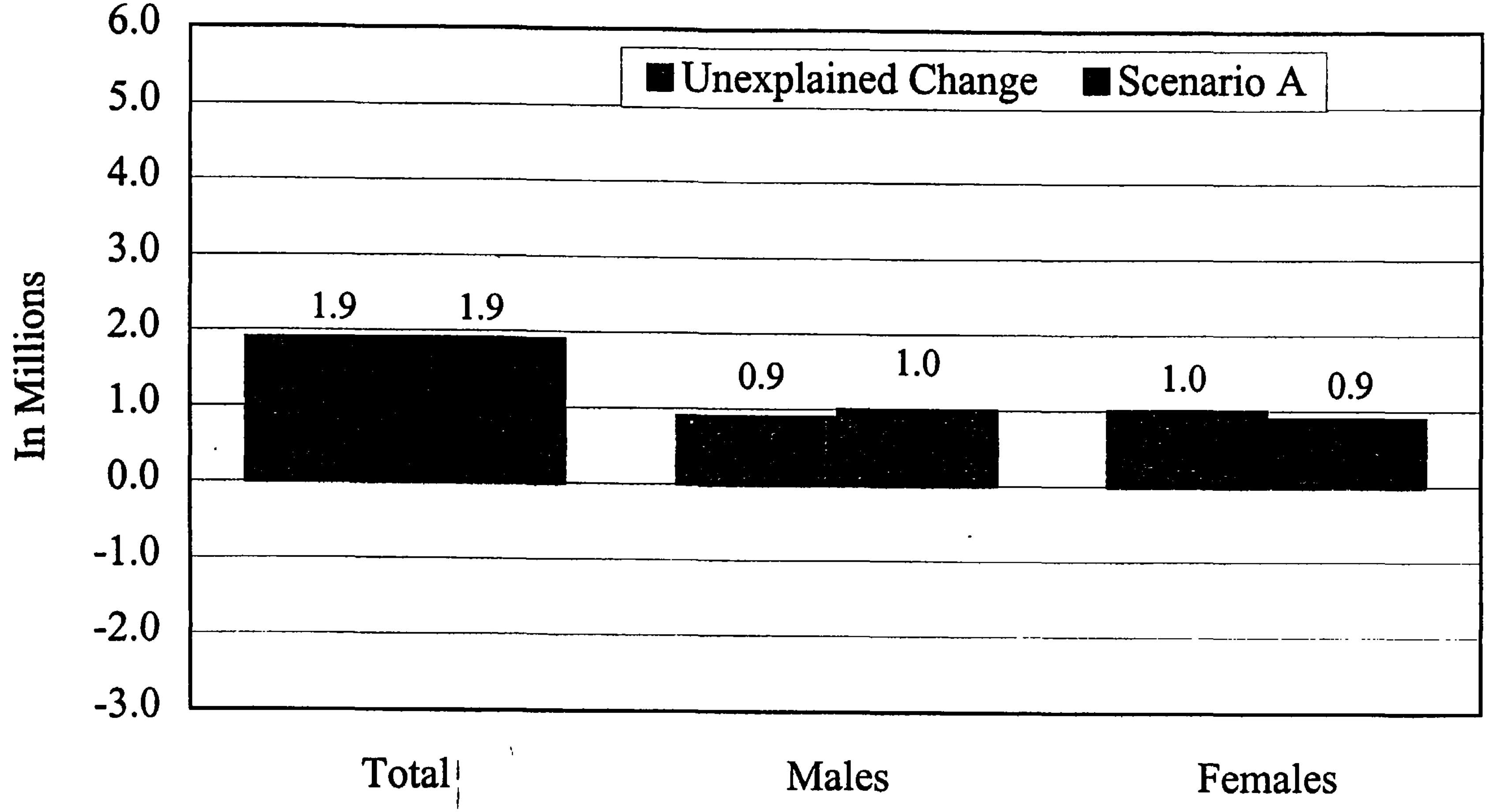
Comparison of Unexplained Change with the Results of Scenario A by Race and Hispanic Origin



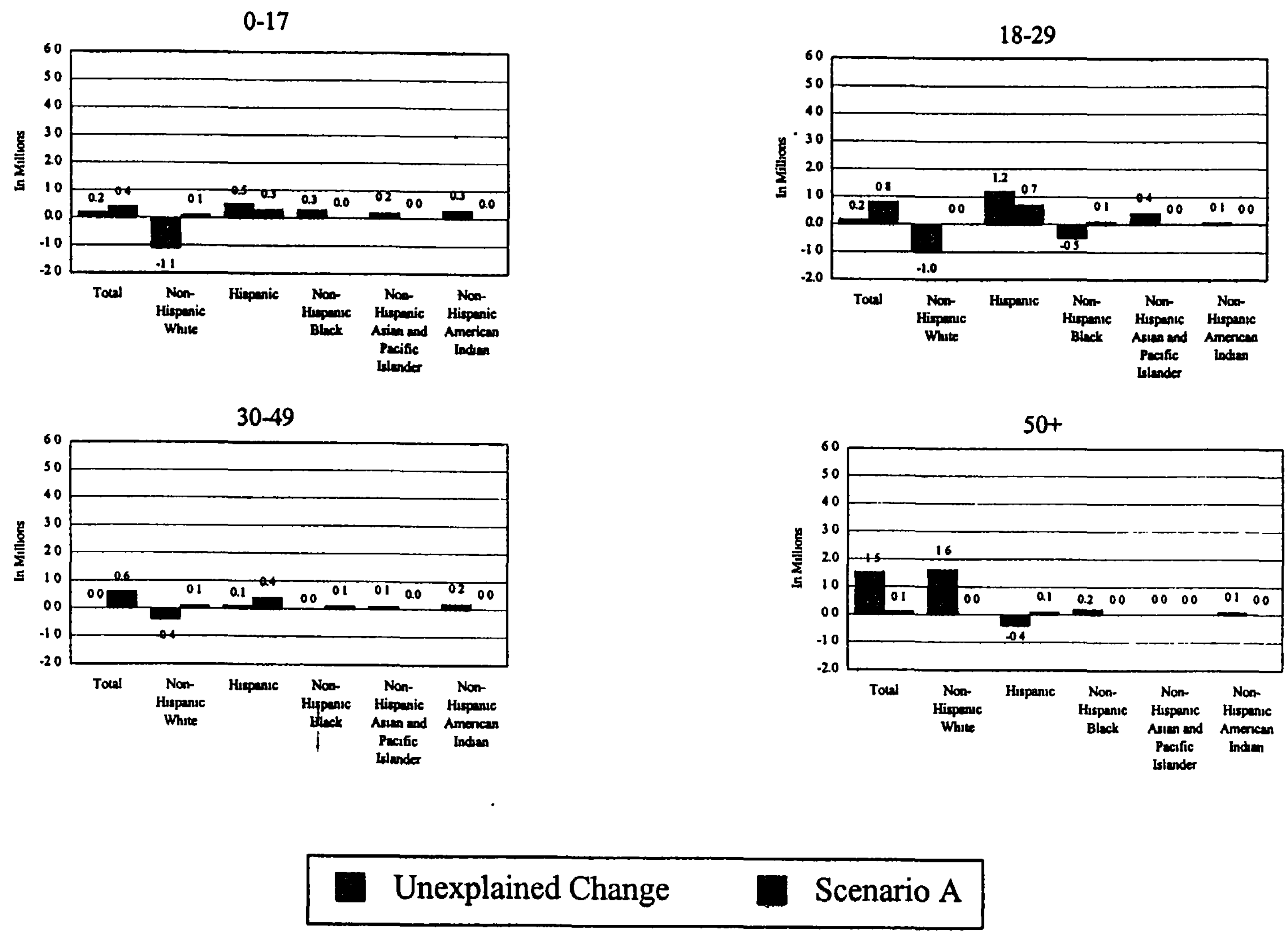
Comparison of Unexplained Change with the Results of Scenario A by Age



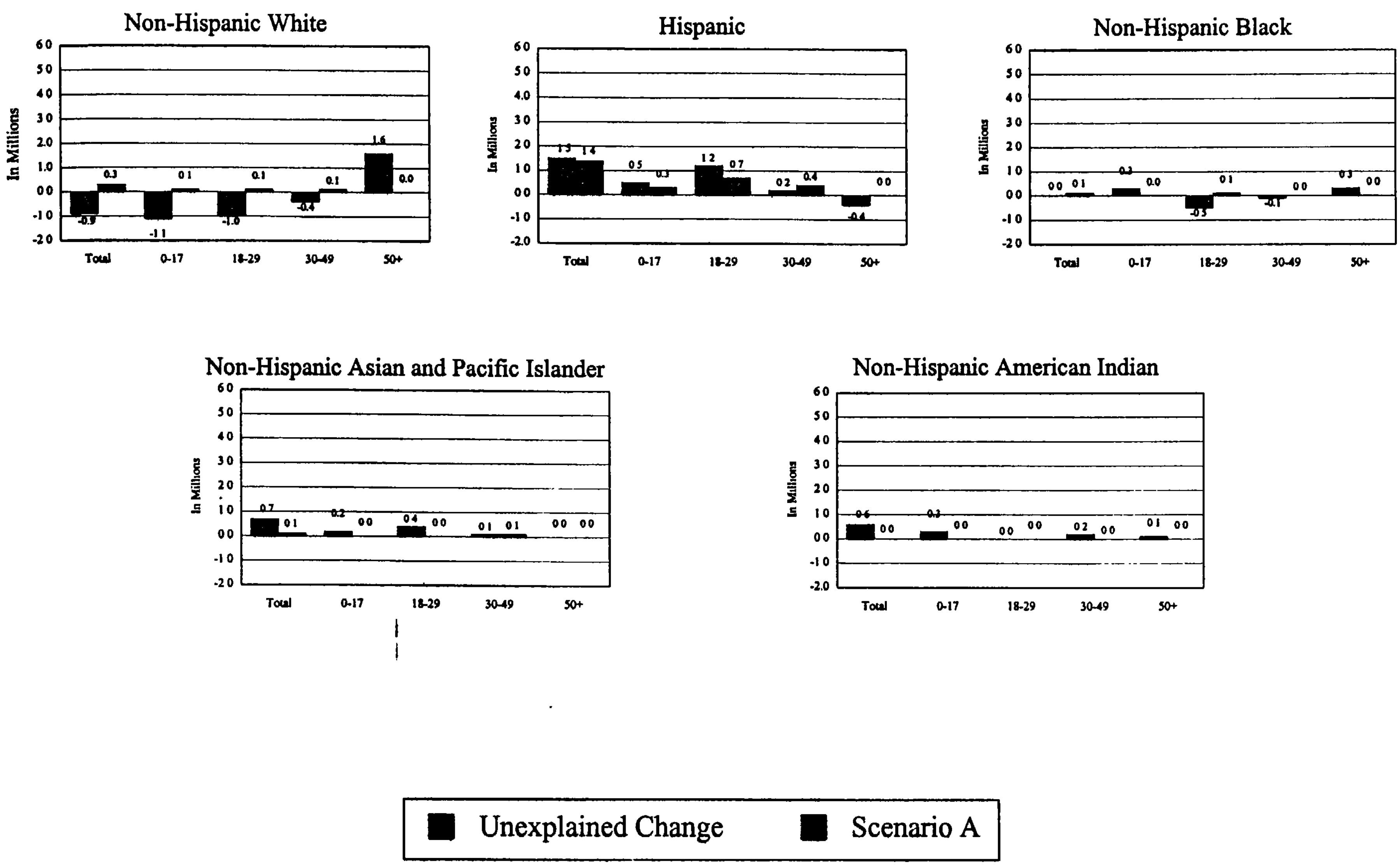
Comparison of Unexplained Change with the Results of Scenario A by Sex



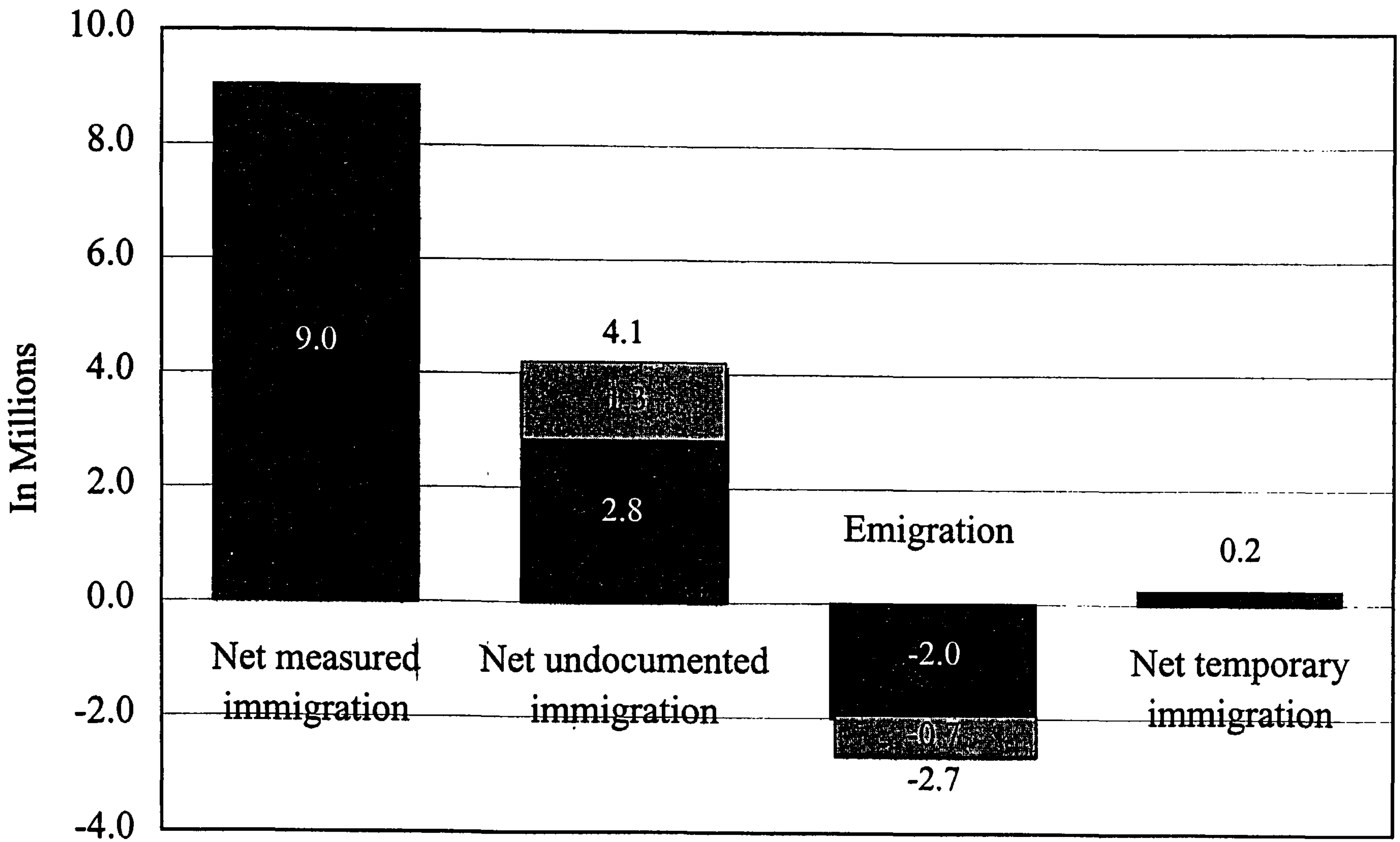
Scenario A by Race and Hispanic Origin for all Age Groups



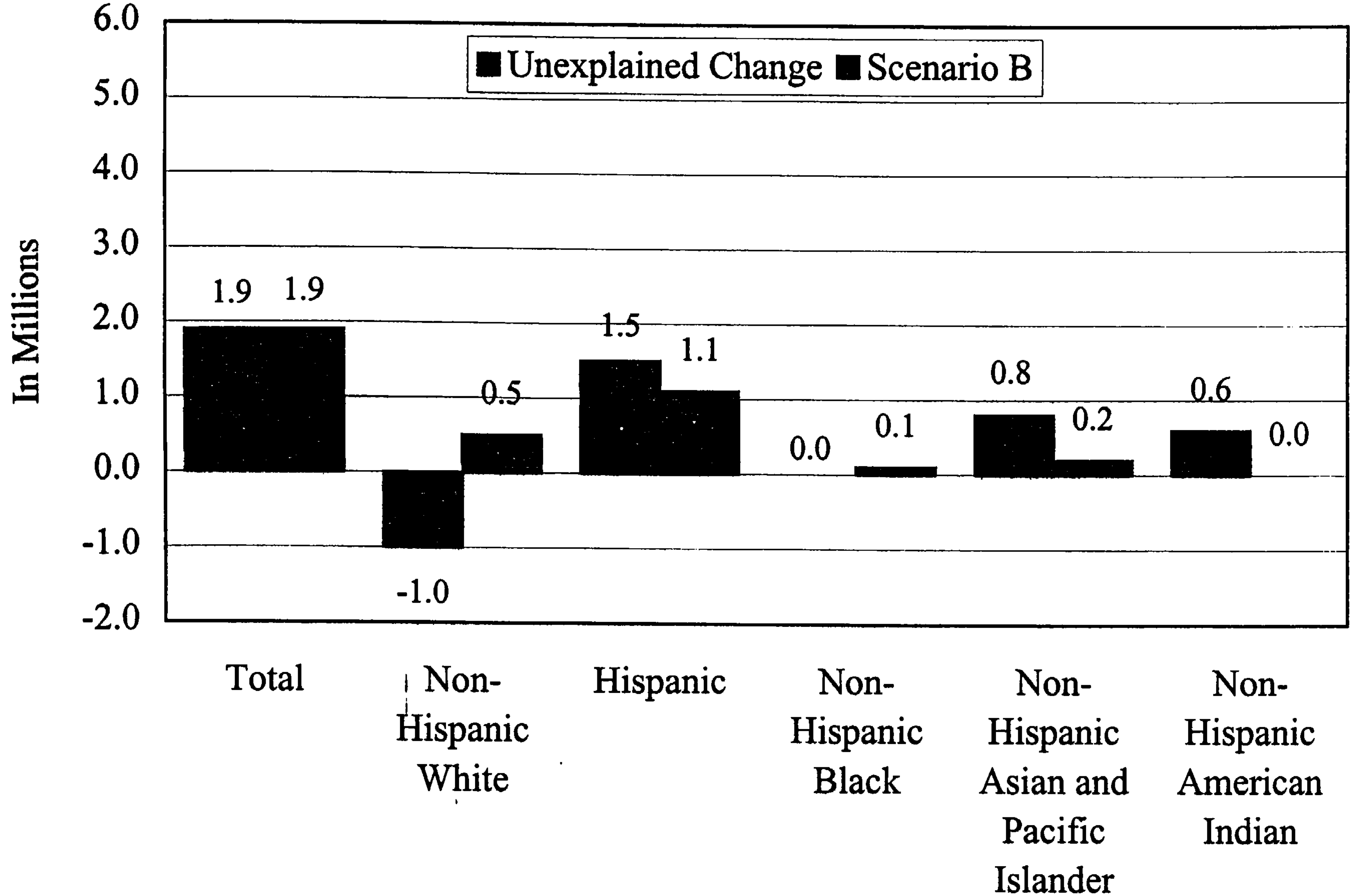
Scenario A by Age for all Race and Hispanic Origin Groups



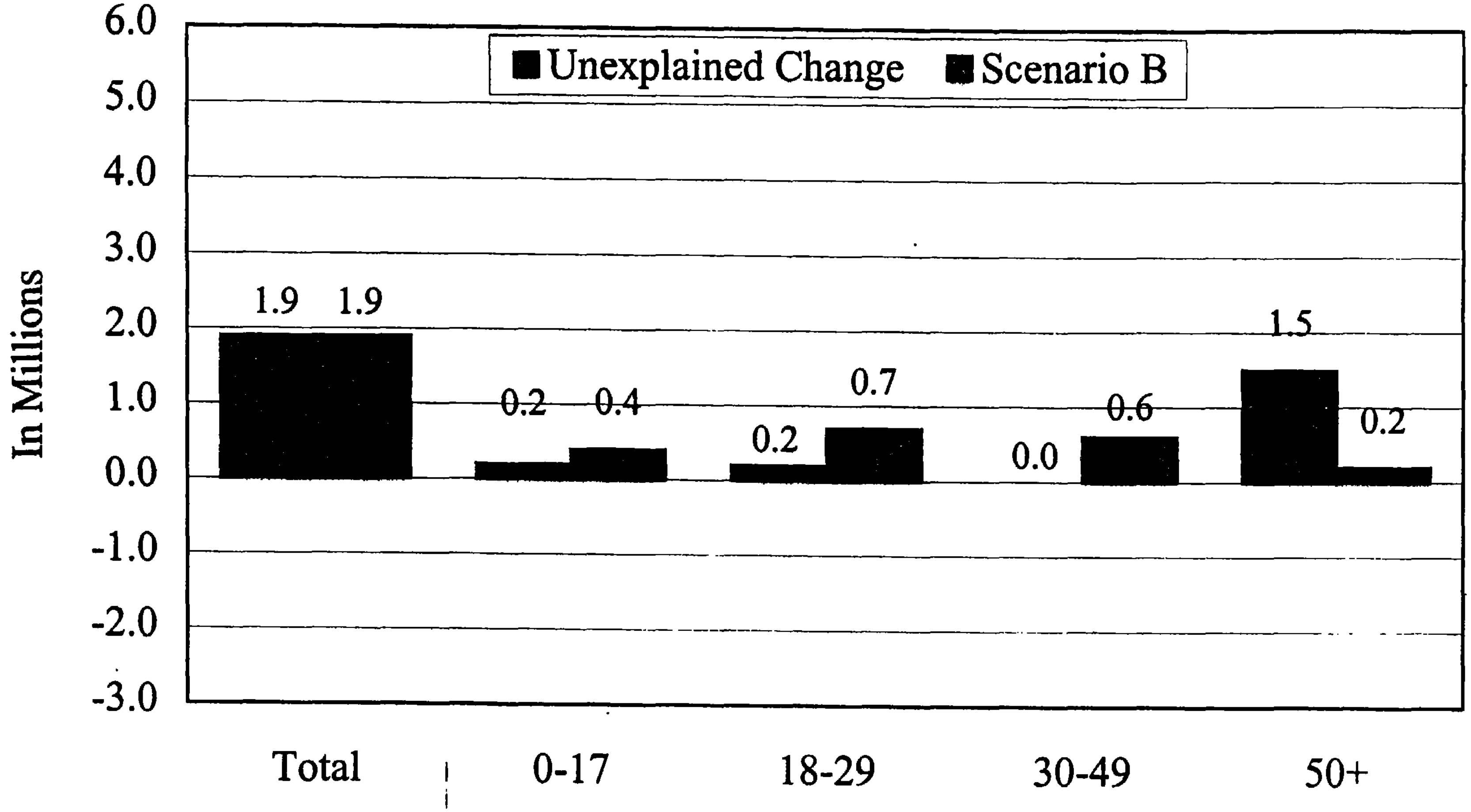
Scenario B: Reduce Emigration by 25%, Remainder Undocumented Migration (1990 Adj. to Census 2000)



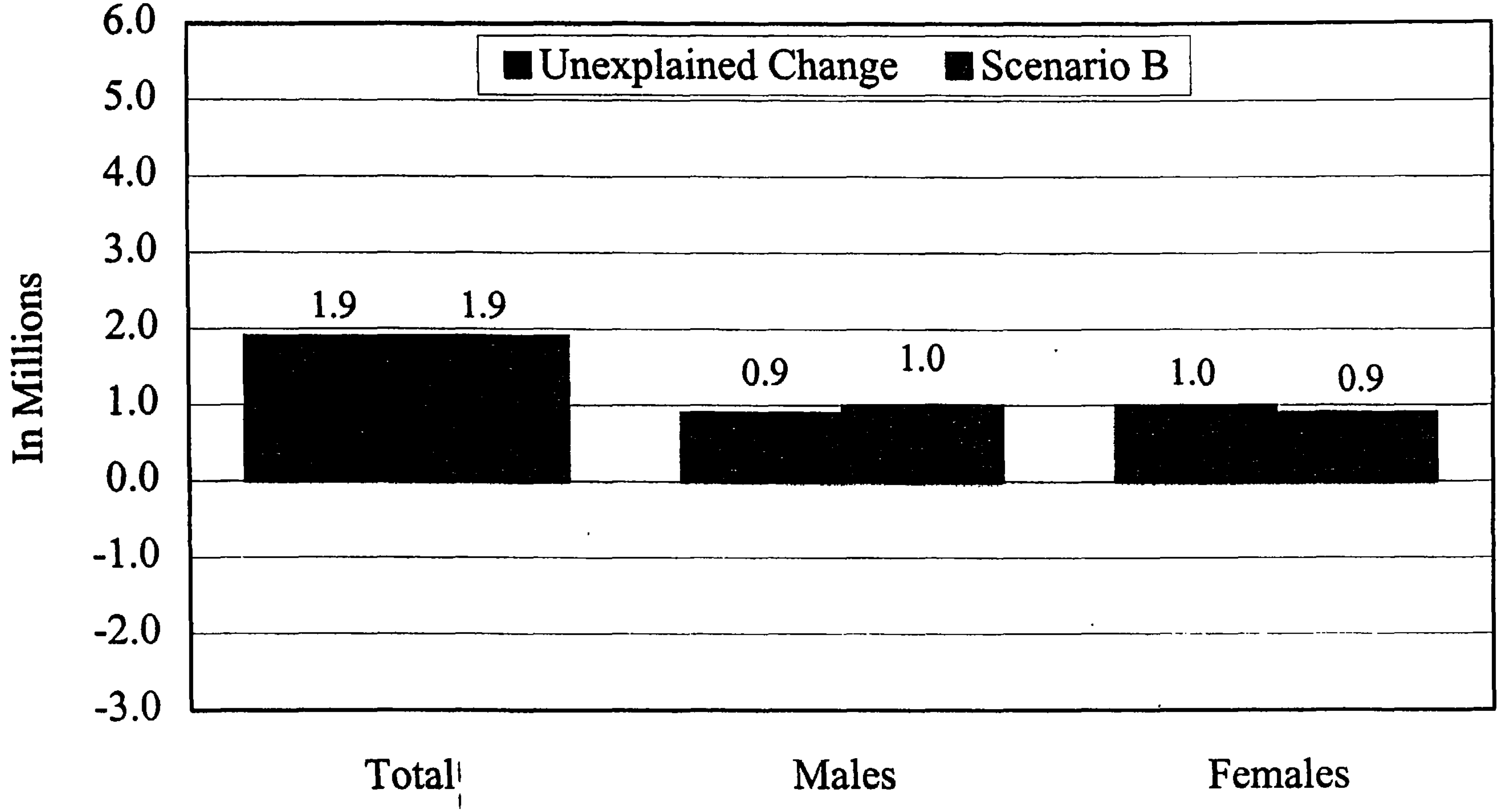
Comparison of Unexplained Change with the Results of Scenario B by Race and Hispanic Origin



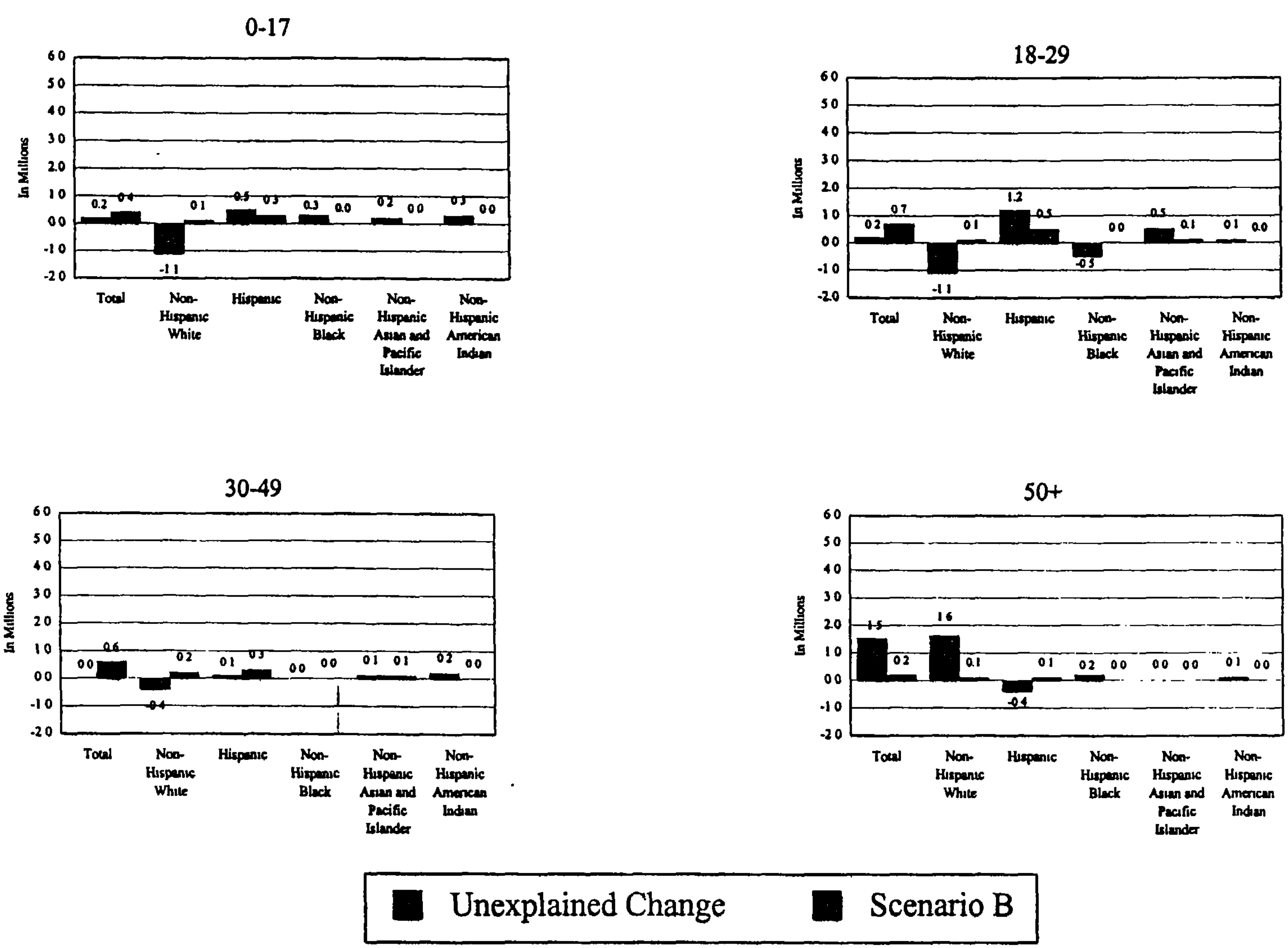
Comparison of Unexplained Change with the Results of Scenario B by Age



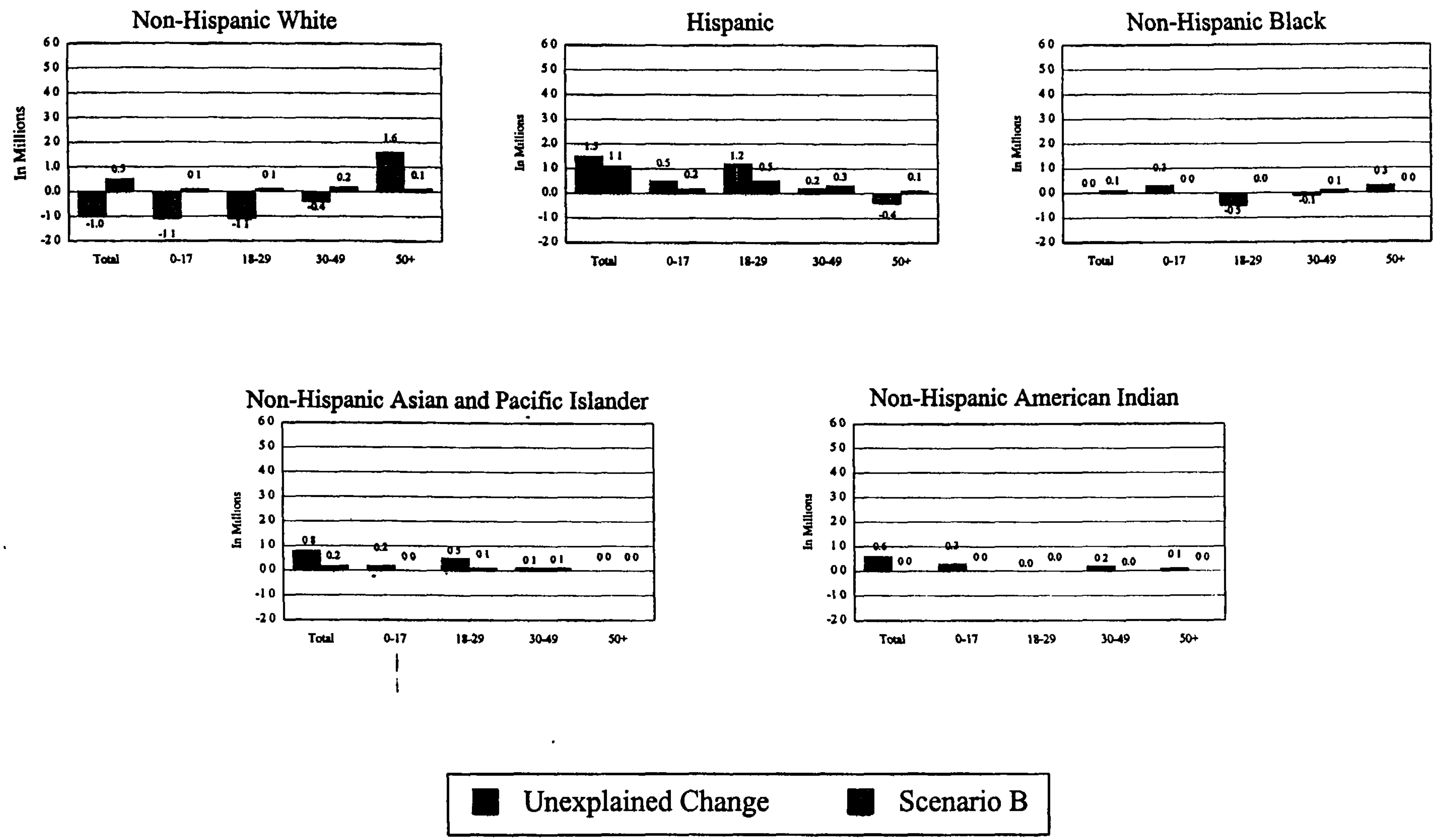
Comparison of Unexplained Change with the Results of Scenario B by Sex



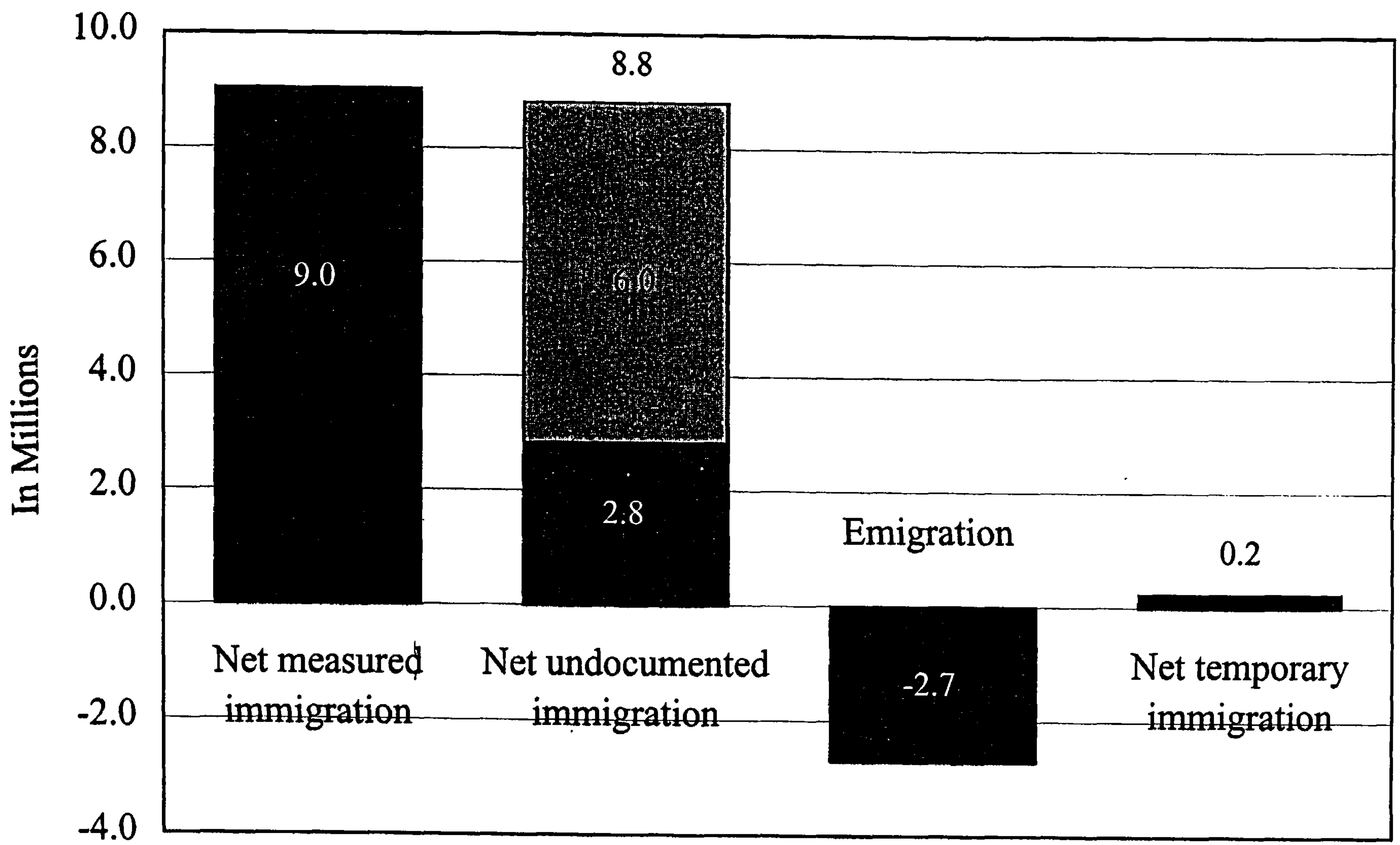
Scenario B by Race and Hispanic Origin for all Age Groups



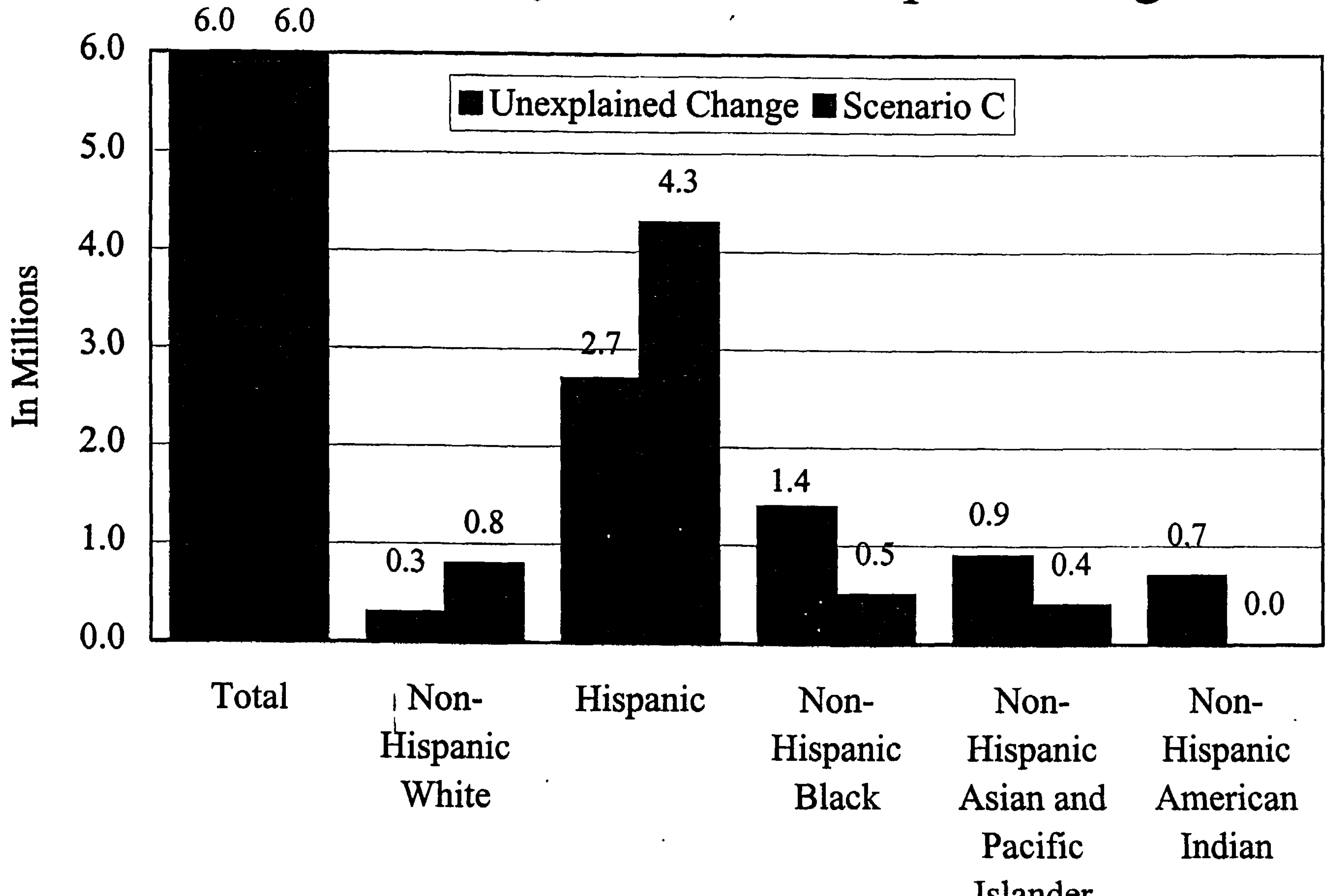
Scenario B by Age for all Race and Hispanic Origin Groups



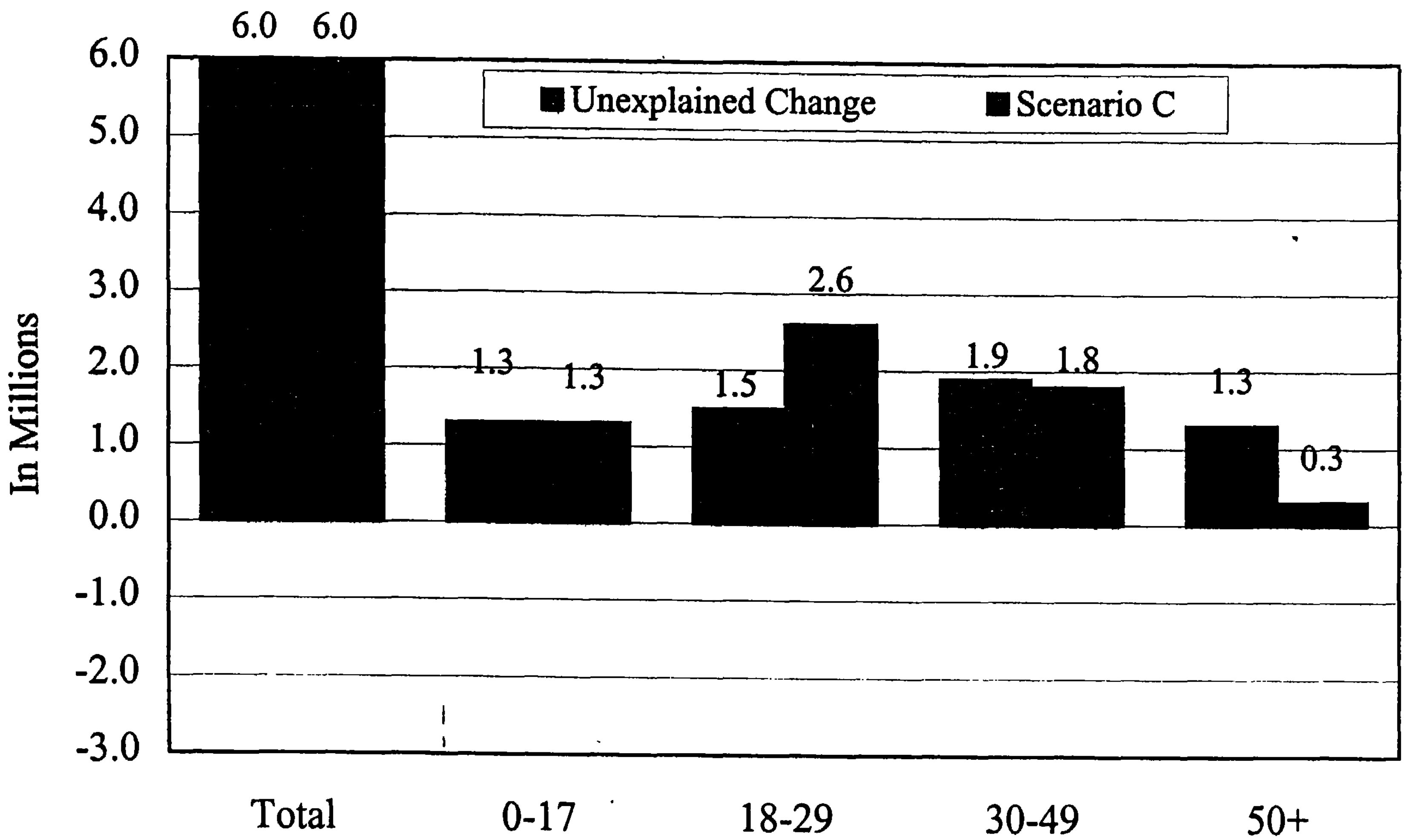
Scenario C: Increase Undocumented Migration (1990 Census to Census 2000)



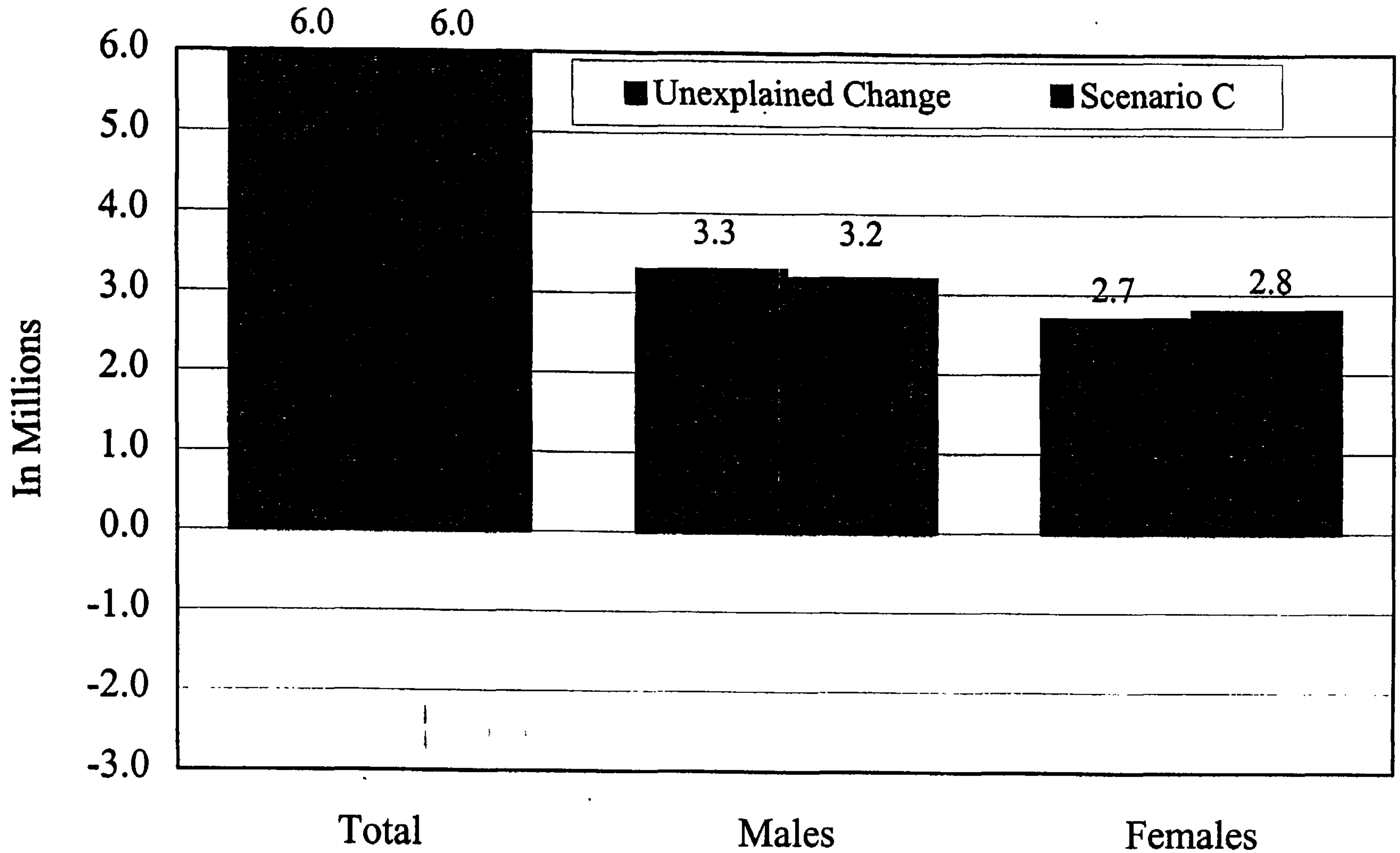
Comparison of Unexplained Change with the Results of Scenario C by Race and Hispanic Origin



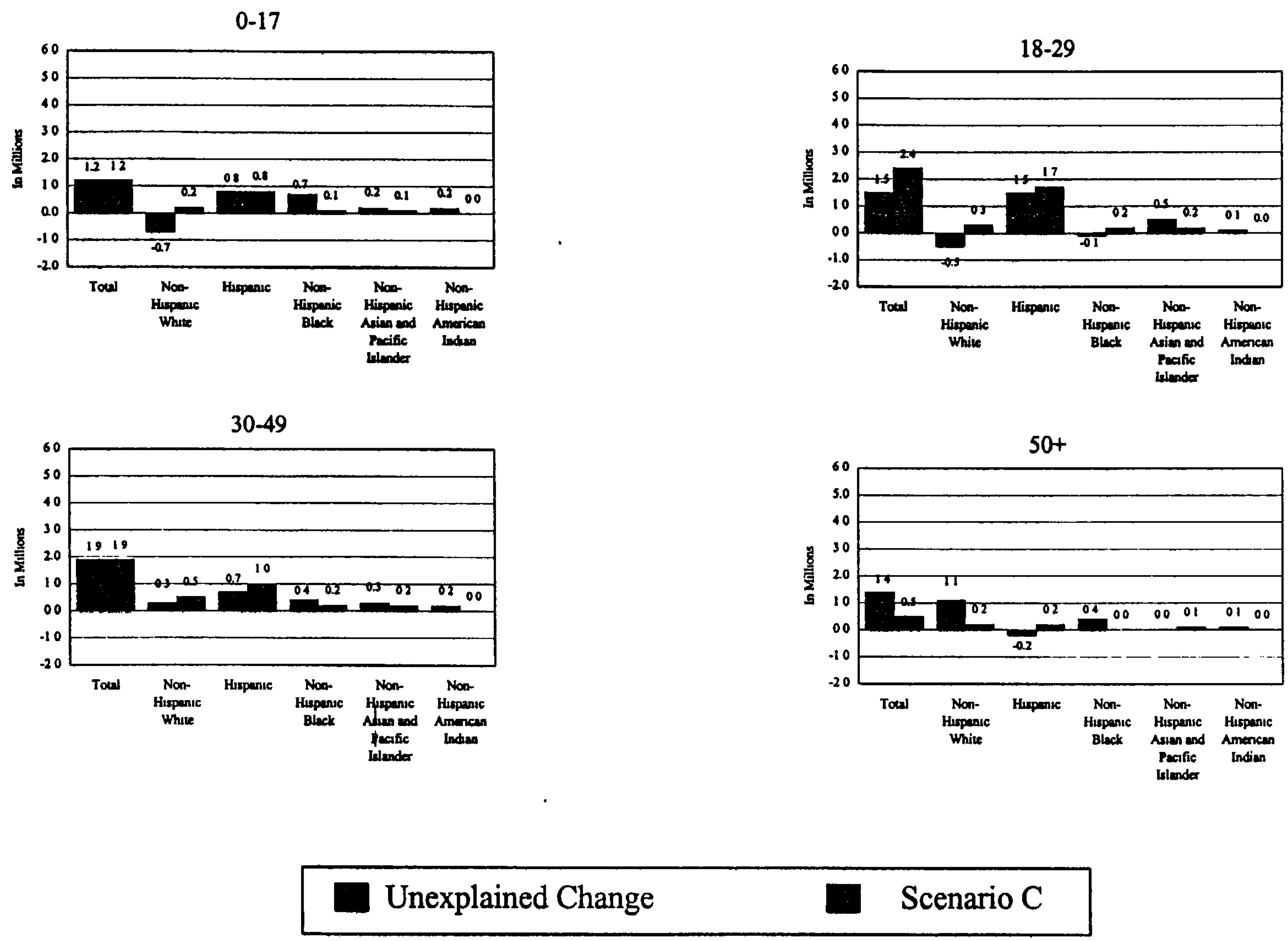
Comparison of Unexplained Change with the Results of Scenario C by Age



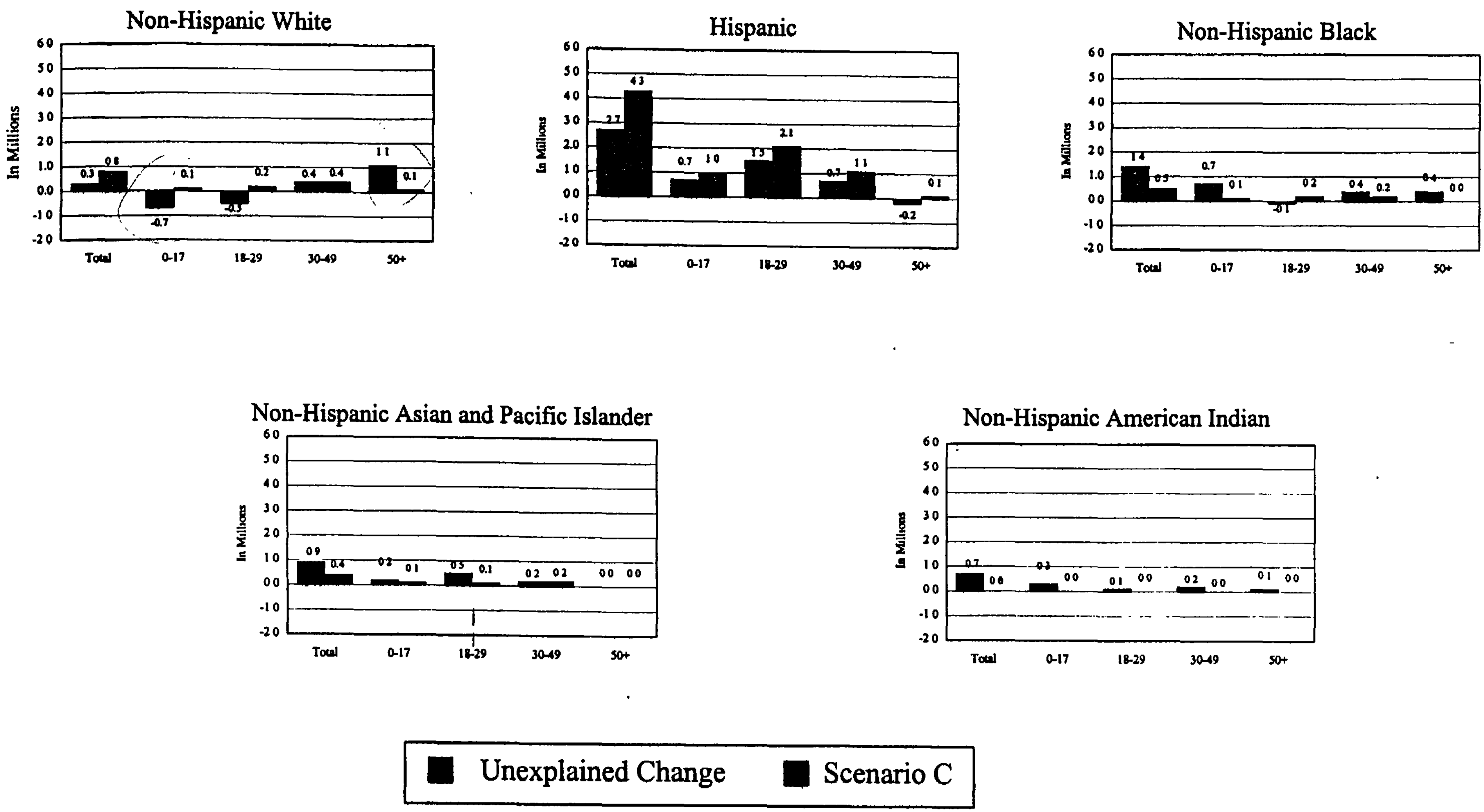
Comparison of Unexplained Change with the Results of Scenario C by Sex



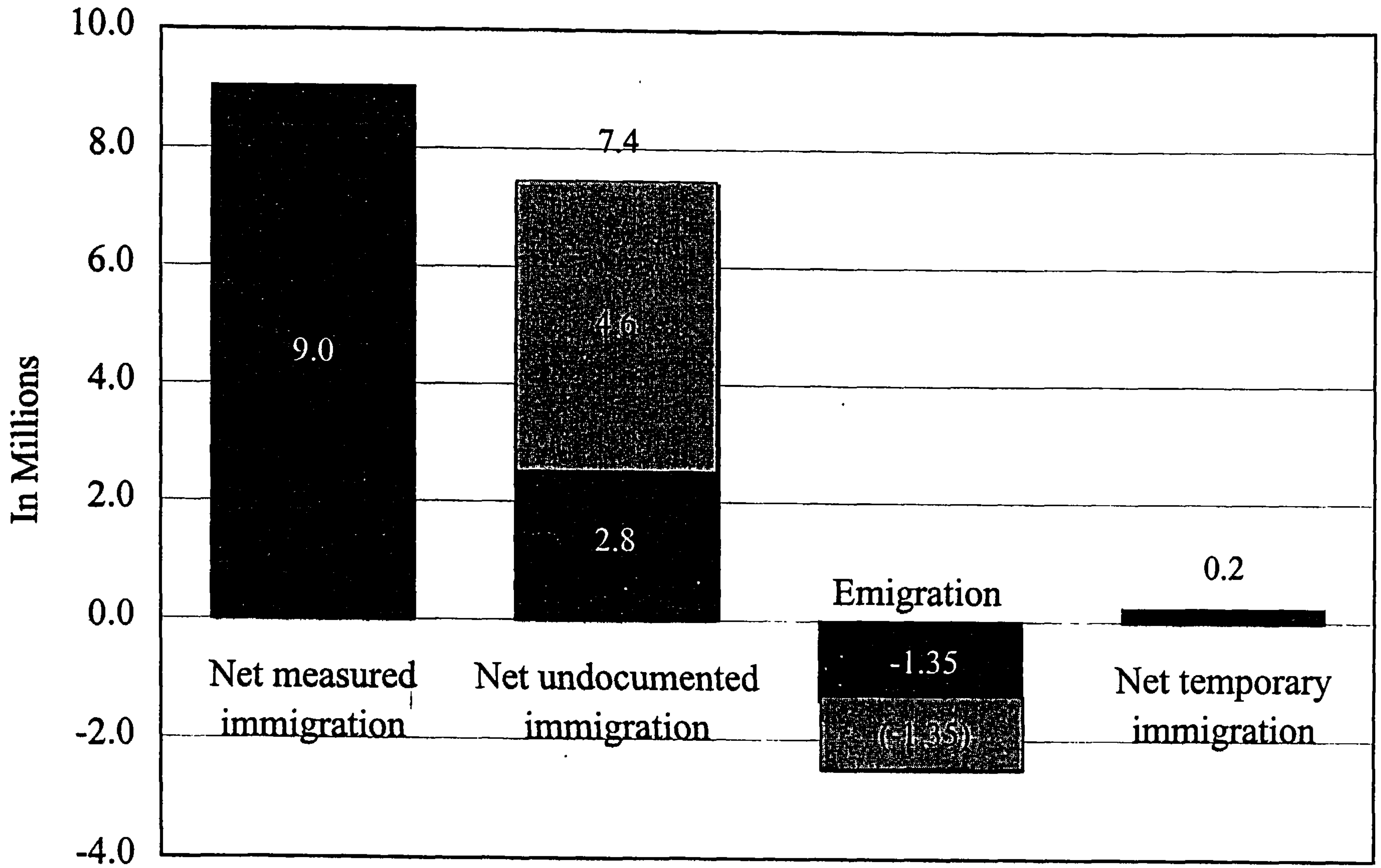
Scenario C by Race and Hispanic Origin for all Age Groups



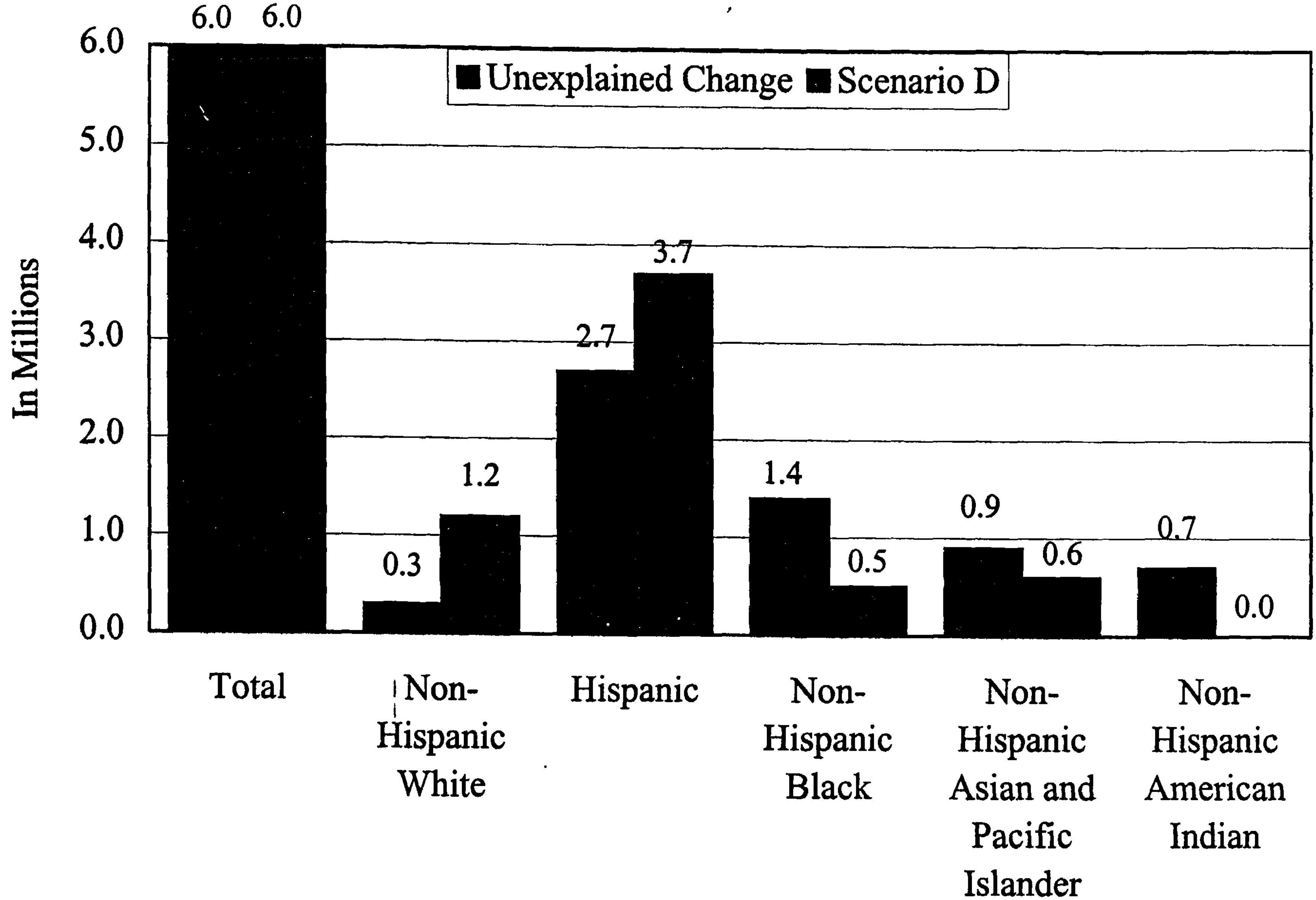
Scenario C by Age for all Race and Hispanic Origin Groups



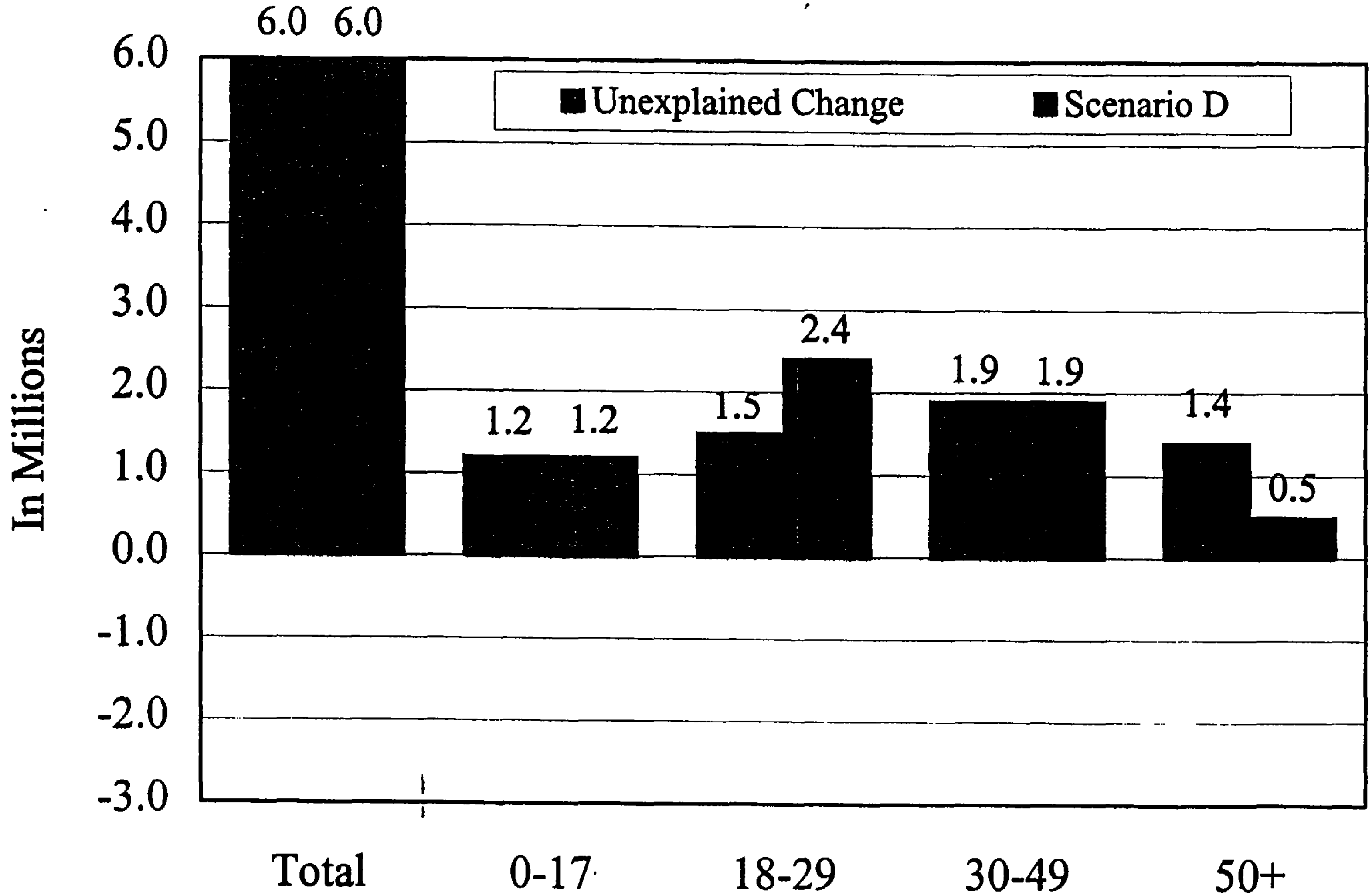
Scenario D: Half Emigration, Remainder Undocumented Migration (1990 Census to Census 2000)



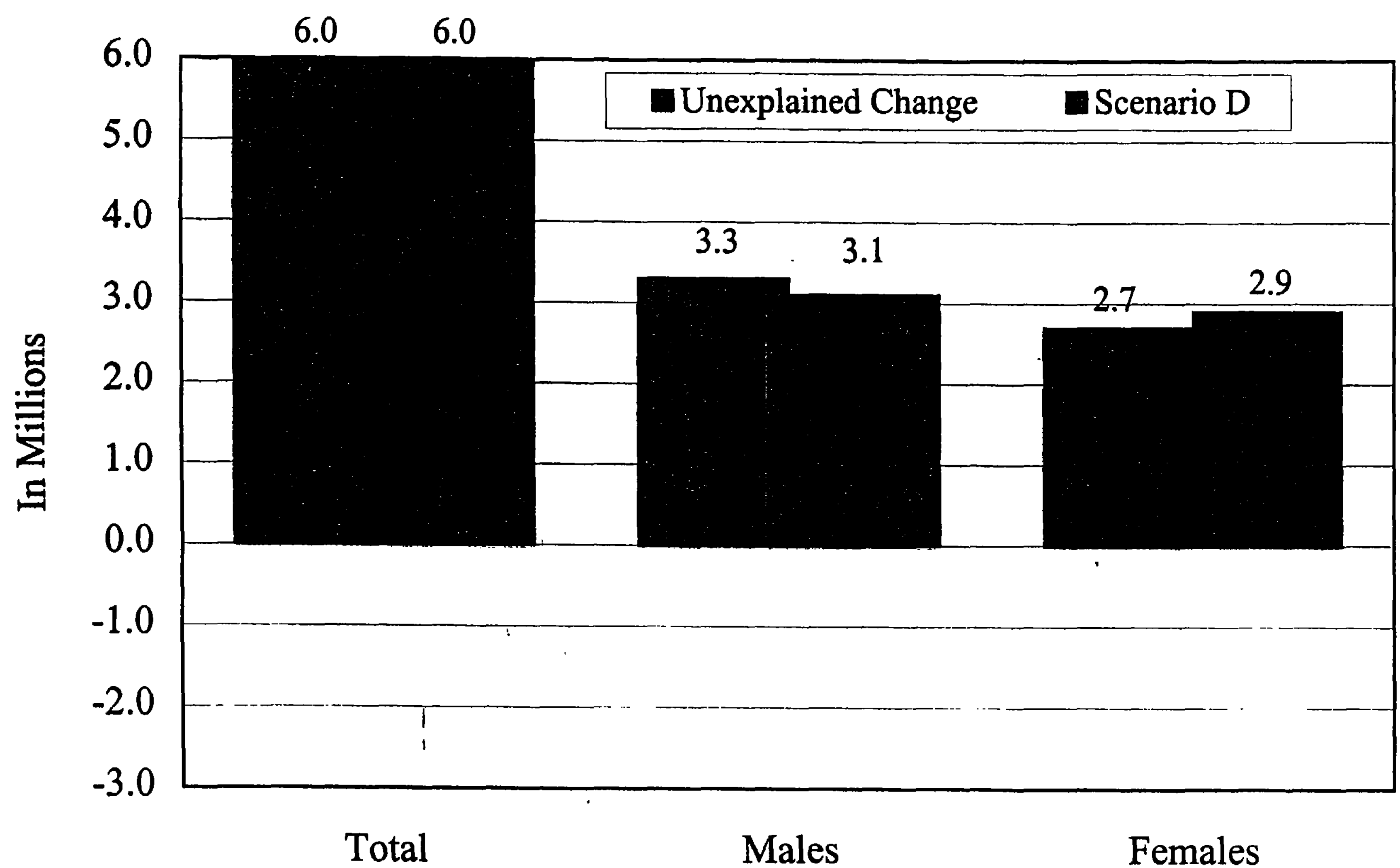
Comparison of Unexplained Change with the Results of Scenario D by Race and Hispanic Origin



Comparison of Unexplained Change with the Results of Scenario D by Age



Comparison of Unexplained Change with the Results of Scenario D by Sex



Alternate Scenario D by Age for all Race and Hispanic Origin Groups

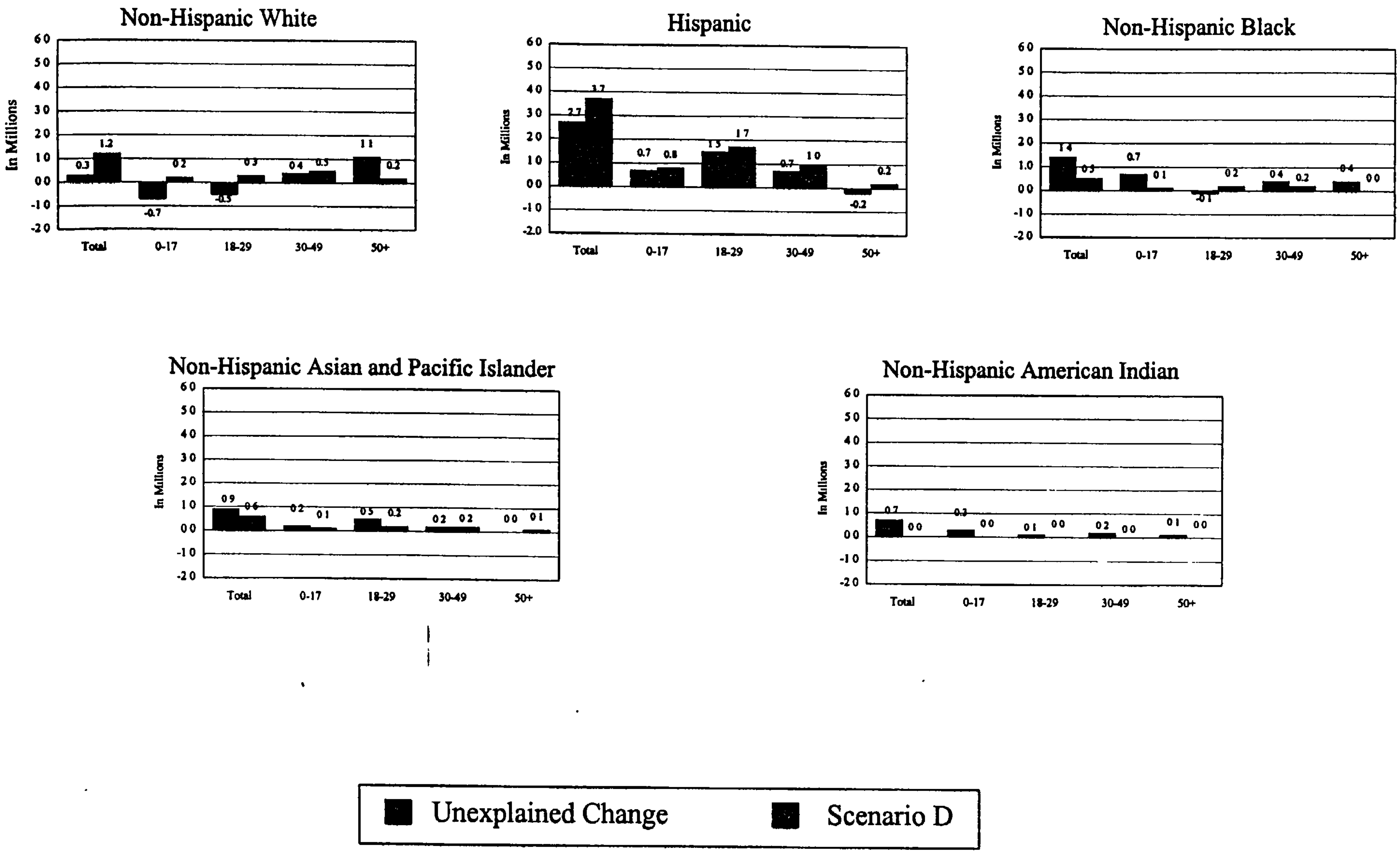


Table 1

INFORMATION DELETED

ESCAP MEETING NO. 43 - 02/21/01

MINUTES

Minutes of the Executive Steering Committee on Accuracy and Coverage Evaluation (A.C.E.) Policy (ESCAP) Meeting # 43

February 21, 2001

Prepared by: Sarah Brady

The forty-third meeting of the Executive Steering Committee on Accuracy and Coverage Evaluation Policy was held on February 21, 2001 at 2:30. The agenda for the meeting was to discuss the treatment of movers in the A.C.E., results from demographic analysis scenarios, distribution of late adds and imputed cases by region and post-stratum.

Committee Attendees:

Nancy Potok
Paula Schneider
Cynthia Clark
Nancy Gordon
John Thompson
Jay Waite
Bob Fay
Howard Hogan
Ruth Ann Killion
John Long
Carol Van Horn

Deputy Director/Acting Director:
William Barron

Other Attendees:

Marvin Raines
William Bell
Kathleen Styles
Maria Urrutia
Sarah Brady

I. Treatment of Movers in A.C.E.

Howard Hogan described the treatment of movers in the A.C.E. as compared to a different methodology implemented for the 1990 PES. This was presented so the Committee could understand the possible effect of the change on the total error model. The committee concluded that the treatment of movers in the A.C.E. improved the matching error relative to 1990, but potentially increased the correlation bias relative to 1990. Therefore, the level of matching error used in the total error model is a conservative assumption regarding the effect of movers. In addition, the potential increase in correlation bias illustrates why a thorough analysis of the sensitivity of the loss functions to correlation bias is important.

II. Demographic Analysis Scenarios

John Long presented alternative scenarios to explain the difference between demographic analysis results and the A.C.E. The scenarios attempted to explain the difference by (1) comparing the 1990 adjusted census to the 2000 unadjusted census and (2) comparing the 1990 unadjusted census to the 2000 unadjusted census. For (1) the difference was examined by two different scenarios-all of the difference was due to net undocumented immigration or a 25 percent reduction in emigration with the remaining difference in net undocumented immigration. The difference described in (2) was examined by two different scenarios-all of the difference was due to net undocumented immigration or a 50 percent reduction in emigration with the remaining difference in net undocumented immigration. The attached documents describes each of the scenarios and presents the results.

Overall, the Committee concluded that the difference between the A.C.E. and DA could not be explained satisfactorily by strictly an increase in undocumented immigration. Additional DA research will be presented at tomorrow's meeting.

John Long also provided data that was requested by Jay Waite comparing the proportion of college aged people in housing units and in dorms for 1990 and 2000. The data had been requested to see if there was evidence that the A.C.E. had not measured duplication between the group quarters population and the housing unit population for this age group. These data are attached. The committee noted that there was a minimal difference between 1990 and 2000. Consequently, this issue is no longer a concern.

III. Distribution of Late Census Adds and Whole Person Imputations (IIs)

Howard Hogan distributed data for the distribution of late census adds and IIs for post-stratum groups by region. The Committee noted that the distribution for IIs appeared to more consistently distributed than the distribution for late adds. The distribution of late census adds was very clustered in some regions for certain post-stratum groups. The Committee concluded that this could possibly affect the synthetic assumption, again raising concerns that this must be studied**.**

IV. Next Meeting

The agenda for the next meeting, scheduled for February 22, 2001 is to discuss unresolved issues and concerns and to present results from the revised demographic analysis.

ESCAP MEETING NO. 44 - 02/22/01

AGENDA

There was no agenda developed or used for the February 22, 2001 meeting.

ESCAP MEETING NO. 44 - 02/22/01

HANDOUTS

Materials attached to these minutes were draft and preliminary material to inform the ESCAP Committee. The data and analysis contained in these documents are subject to revision and are not final. These materials report the results of research and analysis undertaken by Census Bureau staff. They have undergone a more limited review than official Census Bureau publications. Research results and conclusions expressed are those of the authors and do not necessarily indicate concurrence by the Census Bureau.

Presentation on Demographic Analysis

by

J. Gregory Robinson

February 22, 2001

Revised 2-21-01

Table 1- Census Count, Demographic Analysis (DA) Estimates, and Accuracy and Coverage Evaluation (A.C.E.) Estimate for the U.S. Resident Population: 4-1-2000

	Count or Estimate
1. Census Count	281,421,906
2. D.A. Estimate	
a. Base Set	279,598,121
b. Alternative Set	282,335,711
3. A.C.E. Estimate	284,683,785
Difference from Census:	
4. D.A. Estimate	
a. Base Set (=2a-1)	(1,823,785)
b. Alternative Set (=2b-1)	913,805
5. A.C.E. Estimate (=3-1)	3,261,879
Percent Difference	
4. D.A. Estimate	
a. Base Set (=4a/2a*100)	-0.65
b. Alternative Set (=4b/2b*100)	0.32
5. A.C.E. Estimate (=5/3*100)	1.15

Note: The DA estimates for ages under 65 are based on components of population change (births, deaths, legal immigration, and estimates of emigration and undocumented immigration). The DA estimates for ages 65 and over are based on 2000 Medicare data, adjusted for underenrollment.

The A.C.E. and DA estimates are preliminary.

D.A. Base Set - DA estimates without alternative assumptions.
D.A. Alternative Set - DA estimates with alternative assumption that doubles the estimated number of undocumented immigrants entering during the 1990's (from 2.75 to 5.5 million).

Revised 2-21-01

Table 2--Estimates of Percent Net Undercount by Sex: 1940 to 2000
(a minus sign denotes a net overcount)

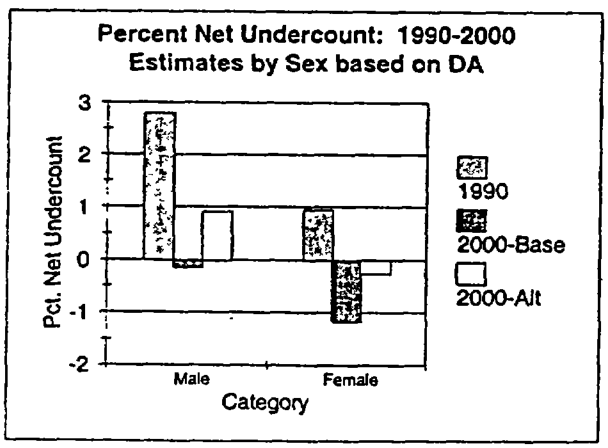
Category	Demographic Analysis								Survey-based	
							2000		PES	A.C.E.
	1940	1950	1960	1970	1980	1990	Base DA	Alt DA	1990	2000
Total Popu	5.4	4.1	3.1	2.7	1.2	1.8	-0.7	0.3	1.6	1.15
Male	5.8	4.4	3.5	3.4	2.2	2.8	-0.1	0.9	1.9	1.5
Female	5.0	3.8	2.7	2.0	0.3	0.9	-1.2	-0.2	1.3	0.8
Male:Fem	0.8	0.6	0.8	1.4	1.9	1.9	1.0	1.2	0.7	0.7

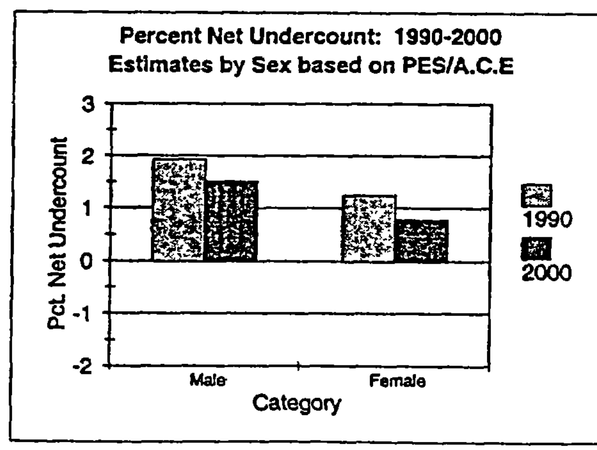
Source: 1940-1990-- Robinson, J. Gregory, Bashir Ahmed, Prithwis Das Gupta, and Karen Woodrow, "Estimates of Population Coverage in the 1990 United States Census Based on Demographic Analysis", Journal of the American Statistical Association, Vol. 88, No. 423, pp. 1061-1077. Estimates for 2000 are unpublished preliminary results.

Source: 2000 - Preliminary A.C.E. and DA estimates. Universe is the U.S. resident population.

Note: D.A. Base Set - DA estimates without alternative assumptions.
D.A. Alternative Set - DA estimates with alternative assumption that doubles the estimated number of undocumented immigrants entering during the 1990's (from 2.75 to 5.5 million).

Figure 1





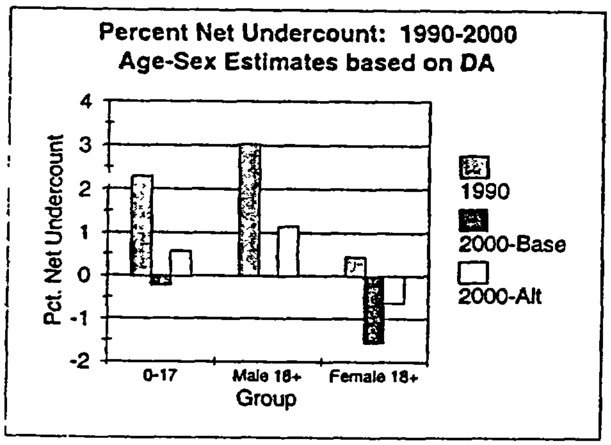
Revised 2-21-01

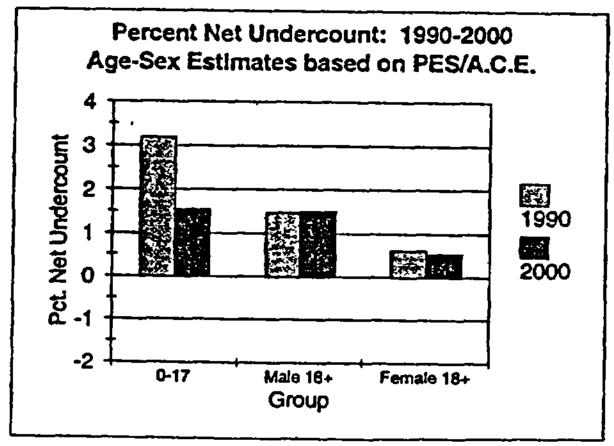
Table 3--Estimates of Percent Net Undercount by Sex and Age: 1960 to 2000
(a minus sign denotes a net overcount)

	Demographic Analysis						Survey-based	
					2000		PES	A.C.E.
Category	1960	1970	1980	1990	Base DA	Alt DA	1990	2000
MALE								
Total	3.5	3.4	2.2	2.8	-0.1	0.9	1.9	1.5
0-17	2.8	2.7	0.9	2.2	-0.5	0.3	3.2	1.5
18-29	5.9	3.9	3.3	2.2	-2.5	0.3	3.2	3.5
30-49	4.2	5.1	3.6	3.8	1.3	2.3	1.9	1.8
50+	2.2	2.5	1.2	2.7	0.2	0.3	-0.6	-0.2
FEMALE								
Total	2.7	2.0	0.3	0.9	-1.2	-0.2	1.3	0.8
0-17	1.8	2.4	0.9	2.4	0.1	0.9	3.2	1.5
18-29	2.8	1.3	0.4	0.6	-3.1	-0.7	2.8	2.1
30-49	1.9	1.3	-0.0	0.5	-0.9	0.0	0.9	1.0
50+	4.6	2.6	-0.2	0.2	-1.4	-1.3	-1.2	-0.8

Note: DA estimates are consistent with estimates in Table 2.

Figure 2





Revised 2-21-01

Table 4--Estimates of Percent Net Undercount by Race and Sex: 1940 to 2000
(a minus sign denotes a net overcount)

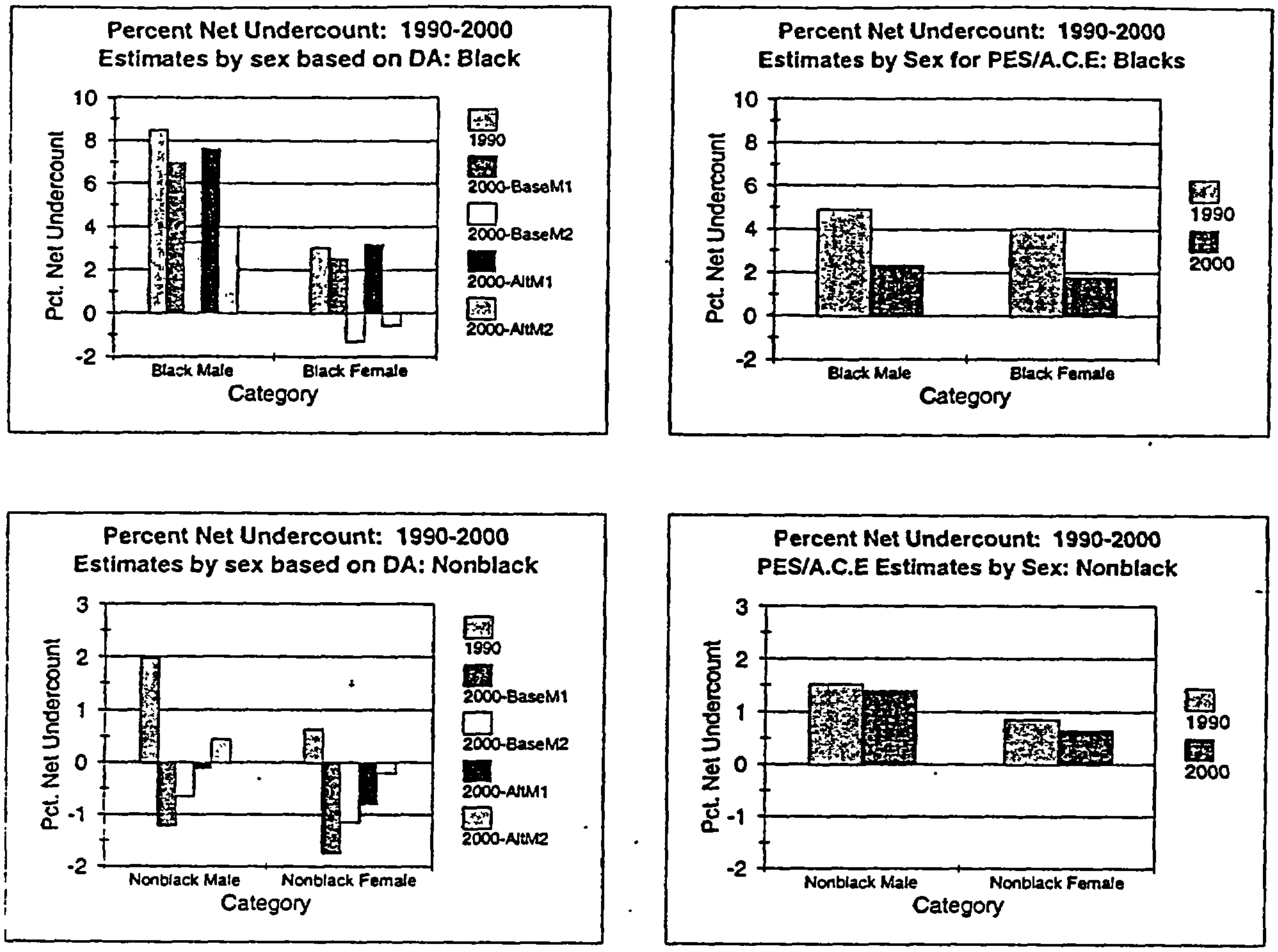
	Demographic Analysis								Survey-based	
					2000-Base DA		2000-Alt DA		PES	A.C.E.
Category	1960	1970	1980	1990	Model 1	Model 2	Model 1	Model 2	1990	2000
Total Popu	3.1	2.7	1.2	1.8	**-0.7**	**-0.7**	**0.3**	**0.3**	1.6	**1.15**
Black	6.6	6.5	4.5	5.7	**4.7**	**0.9**	**5.4**	**1.7**	4.4	**2.1**
Male	8.8	9.1	7.5	8.5	**6.9**	**3.2**	**7.6**	**4.0**	4.9	**2.4**
Female	4.4	4.0	1.7	3.0	**2.5**	**-1.3**	**3.2**	**-0.6**	4.0	**1.8**
Nonblack	2.7	2.2	0.8	1.3	**-1.5**	**-0.9**	**-0.5**	**0.1**	1.2	**1.0**
Male	2.9	2.7	1.5	2.0	**-1.2**	**-0.7**	**-0.1**	**0.4**	1.5	**1.4**
Female	2.4	1.7	0.1	0.6	**-1.7**	**-1.1**	**-0.8**	**-0.2**	0.9	**0.6**
Black:Non	3.9	4.3	3.7	4.4	**6.2**	**1.8**	**5.8**	**1.5**	3.3	**1.0**

Note: Model 1 census tabulations for Blacks include persons who marked the Black circle and no other race response cirlces.
Model 2 census tabulations for Blacks include persons who marked the Black circle and other response circles.
Persons who marked only the "Other race" circle are reassigned to a specific race category (to be consistent with 1990 DA estimates and the historical demographic data series)

Source: 1940-1990-- Robinson, J. Gregory, Bashir Ahmed, Prithwis Das Gupta, and Karen Woodrow, "Estimates of Population Coverage in the 1990 United States Census Based on Demographic Analysis", Journal of the American Statistical Association, Vol. 88, No. 423, pp. 1061-1077. Estimates for 2000 are unpublished preliminary results.

Source: 2000 - See Table 2. Note that the A.C.E. estimates for Blacks pertain to the Non-Hispanic Blacks.

Figure 3



Percent Net Undercount: 1990-2000
Estimates by sex based on DA: Black
10
8
6
4
2
0
-2
Pct. Net Undercount
Black Male
Black Female
Category
1990
2000-BaseM1
2000-BaseM2
2000-AltM1
2000-AltM2
Percent Net Undercount: 1990-2000
Estimates by Sex for PES/A.C.E: Blacks
10
8
6
4
2
0
-2
Pct. Net Undercount
Black Male
Black Female
Category
1990
2000
Percent Net Undercount: 1990-2000
Estimates by sex based on DA: Nonblack
3
2
1
0
-1
-2
Pct. Net Undercount
Nonblack Male
Nonblack Female
Category
1990
2000-BaseM1
2000-BaseM2
2000-AltM1
2000-AltM2
Percent Net Undercount: 1990-2000
PES/A.C.E Estimates by Sex: Nonblack
3
2
1
0
-1
-2
Pct. Net Undercount
Nonblack Male
Nonblack Female
Category
1990
2000

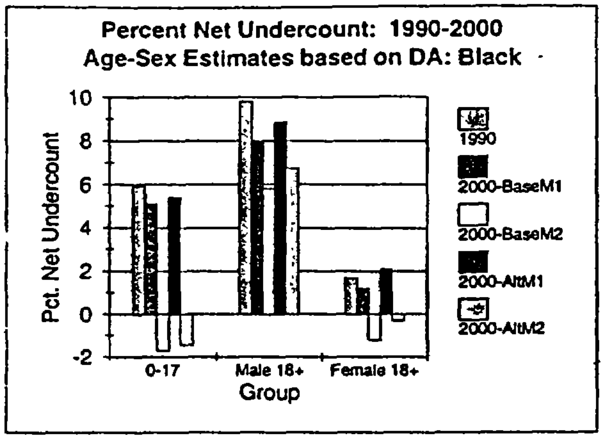
Revised 2-21-01

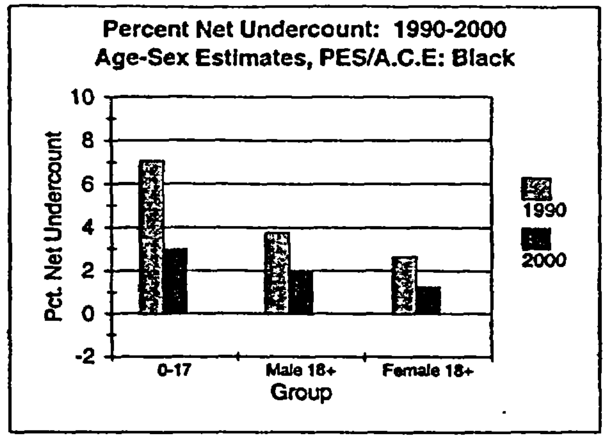
Table 5--Estimates of Percent Net Undercount by Race, Sex and Age: 1960 to 2000
(a minus sign denotes a net overcount)

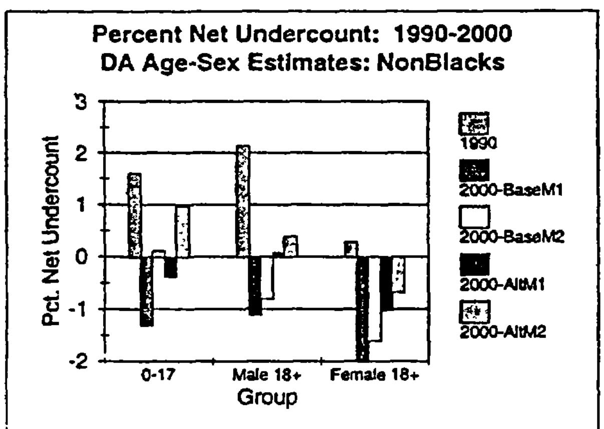
	Demographic Analysis								Survey-based	
					2000-Base DA		2000-Alt DA		PES	A.C.E.
Category	1960	1970	1980	1990	Model 1	Model 2	Model 1	Model 2	1990	2000
BLACK MALE										
Total	8.8	9.1	7.5	8.5	6.9	3.3	7.6	4.0	4.9	2.4
0-17	5.4	6.2	4.2	5.9	4.9	-1.9	5.1	-1.6	7.0	3.0
18-29	15.1	12.1	9.2	7.7	8.0	4.9	9.6	6.5	3.6	3.7
30-49	11.9	14.5	13.1	12.3	10.1	8.3	11.0	9.1	6.3	2.6
50+	6.6	6.3	4.6	8.3	4.1	2.5	4.2	2.6	-0.4	-0.7
BLACK FEMALE										
Total	4.4	4.0	1.7	3.0	2.5	-1.3	3.2	-0.6	4.0	1.8
0-17	4.0	5.6	3.9	5.9	5.4	-1.6	5.7	-1.2	7.1	3.0
18-29	5.4	4.5	2.4	2.9	1.9	-1.7	3.5	-0.0	5.5	3.8
30-49	2.1	0.5	0.6	2.5	2.1	-0.1	2.9	0.8	3.2	1.3
50+	7.6	3.8	-1.9	-0.8	-0.5	-2.2	-0.3	-2.0	-1.2	-0.8
NONBLACK MALE										
Total	2.9	2.7	1.5	2.0	-1.2	-0.7	-0.1	0.4	1.5	1.4
0-17	2.4	2.1	0.3	1.5	-1.6	-0.2	-0.7	0.6	2.5	1.3
18-29	4.6	2.8	2.4	1.3	-4.5	-3.9	-1.3	-0.7	3.1	3.4
30-49	3.4	4.0	2.5	2.7	-0.4	0.2	1.0	1.2	1.3	1.7
50+	1.8	2.2	0.9	2.2	-0.2	-0.1	-0.1	0.1	-0.6	-0.2
NONBLACK FEMALE										
Total	2.4	1.7	0.1	0.6	-1.7	-1.1	-0.8	-0.2	0.9	0.6
0-17	1.5	1.8	0.3	1.8	-1.0	0.4	-0.1	1.3	2.5	1.3
18-29	2.4	0.9	0.1	0.3	-4.0	-3.3	-1.4	-0.8	2.4	1.8
30-49	1.9	1.3	-0.1	0.2	-1.4	-1.0	-0.4	-0.1	0.6	0.9
50+	4.3	2.5	-0.0	0.3	-1.5	-1.4	-1.4	-1.2	-1.2	-0.8

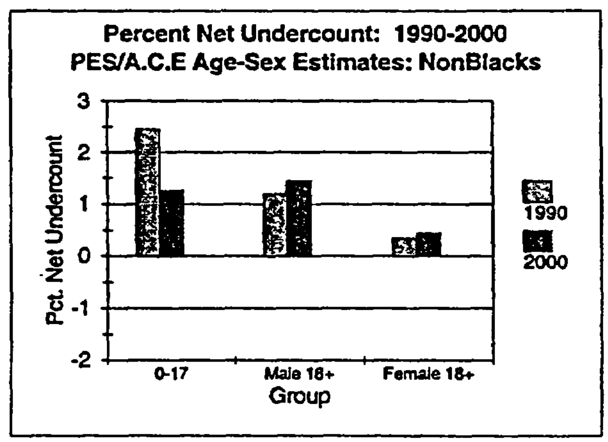
Sources and notes: See Table 2 and 4

Figure 4


Percent Net Undercount: 1990-2000
Age-Sex Estimates based on DA: Black
Pct. Net Undercount
10
8
6
4
2
0
-2
0-17
Male 18+
Female 18+
Group
1990
2000-BaseM1
2000-BaseM2
2000-AltM1
2000-AltM2


Percent Net Undercount: 1990-2000
Age-Sex Estimates, PES/A.C.E: Black
Pct. Net Undercount
10
8
6
4
2
0
-2
0-17
Male 18+
Female 18+
Group
1990
2000


Percent Net Undercount: 1990-2000
DA Age-Sex Estimates: NonBlacks
Pct. Net Undercount
3
2
1
0
-1
-2
0-17
Male 18+
Female 18+
Group
1990
2000-BaseM1
2000-BaseM2
2000-AltM1
2000-AltM2


Percent Net Undercount: 1990-2000
PES/A.C.E Age-Sex Estimates: NonBlacks
Pct. Net Undercount
3
2
1
0
-1
-2
0-17
Male 18+
Female 18+
Group
1990
2000

Materials attached to these minutes were draft and preliminary material to inform the ESCAP Committee. The data and analysis contained in these documents are subject to revision and are not final. These materials report the results of research and analysis undertaken by Census Bureau staff. They have undergone a more limited review than official Census Bureau publications. Research results and conclusions expressed are those of the authors and do not necessarily indicate concurrence by the Census Bureau.

Estimation of Population Coverage in the 1990 United States Census Based on Demographic Analysis

J. GREGORY ROBINSON, BASHIR AHMED, PRITHWIS DAS GUPTA, and KAREN A. WOODROW*

This article presents estimates of net coverage of the national population in the 1990 census, based on the method of demographic analysis. The general techniques of demographic analysis as an analytic tool for coverage measurement are discussed, including use of the demographic accounting equation, data components, and strengths and limitations of the method. Patterns of coverage displayed by the 1990 estimates are described, along with similarities or differences from comparable demographic estimates for previous censuses. The estimated undercount in the 1990 census was 4.7 million, or 1.85%. The undercount of males (2.8%) was higher than for females (.9%), and the undercount of Blacks (5.7%) exceeded the undercount of Non-Blacks (1.3%). Black adult males were estimated to have the highest rate of undercounting of all groups. Race-sex-age patterns of net coverage in the 1990 census were broadly similar to patterns in the 1980 and 1970 censuses. A final section presents the results of the first statistical assessment of the uncertainty in the demographic coverage estimates for 1990.

KEY WORDS: Coverage error; Demographic analysis; Undercount.

1. INTRODUCTION

The general method of demographic analysis as a tool for coverage evaluation is well developed and has been actively used at the Census Bureau to assess the completeness of coverage in every census since 1960. (See Siegel and Zelnik 1966; U.S. Bureau of the Census 1974; and U.S. Bureau of the Census 1988 for the basic demographic evaluations of the 1960, 1970, and 1980 censuses.) Demographic analysis estimates of coverage have become the benchmark by which national differences in coverage for age, sex, and race groups and changes in coverage over time are measured.

The purpose of the demographic analysis evaluation program for 1990 has been twofold: (1) to evaluate the completeness of coverage of population in the 1990 census based on demographic analysis, and (2) to develop a statistically based assessment of the accuracy of those demographic estimates of net coverage. This article reports the results of the demographic estimates of coverage for 1990 and the assessment of the accuracy of the estimates. An important by-product of the demographic program is the historical estimates of coverage provided for every census since 1940. The demographic estimates of net coverage for 1990 were also used to evaluate the overall quality of the national estimates of net coverage based on the 1990 Post-Enumeration Survey (PES). (See Hogan 1992 for a description of the PES.)

Section 2 describes the methodology of the demographic estimates. Section 3 describes the estimates of coverage in the 1990 census based on demographic analysis and compares the estimates with those for previous censuses. Section 4 presents the results of the first-time assessment of uncertainty in the demographic coverage estimates for 1990. Section 5 presents our conclusions and plans for future research.

* J. Gregory Robinson is Chief, Bashir Ahmed is Demographic Statistician, and Prithwis Das Gupta is Mathematical Statistician, Population Analysis and Evaluation Staff, Population Division, U.S. Bureau of the Census, Washington, DC 20233. Karen A. Woodrow is Adjunct Research Associate, Center for Social and Demographic Analysis, State University of New York, Albany, New York 12222. The authors thank the referees and Special Section Editor for their helpful comments and suggestions.

2. THE GENERAL METHOD OF DEMOGRAPHIC ANALYSIS

Estimation of census coverage based on demographic analysis involves developing demographic estimates of the resident population in various categories, such as age-sex-race groups, by combining various sources of administrative and demographic data. The independent population estimates (P) are then compared with the corresponding census counts (C) to yield an estimate of the net census undercount, u, and net undercount rate, r:

$$u = P - C \tag{1}$$

and

$$r = (u/P)*100. \tag{2}$$

Demographic analysis represents a macro-level approach to measuring coverage, where analytic estimates of net undercount are derived by comparing aggregate sets of data or counts. This approach differs fundamentally from the PES, which represents a micro-level approach where estimates of coverage are based on case-by-case matching with census records for a sample of the population.

The particular analytic procedure used to estimate coverage nationally in 1990 for the various demographic subgroups depends primarily on the nature and availability of the required demographic data. Different demographic techniques were used for the populations under age 55, 55–64, and 65 and over; the total population is the sum of these subgroups. Figure 1 summaries the cohort estimation procedure for each group.

2.1 Estimation of Subgroups

2.1.1 Age under 55. The demographic analysis estimates for the population below age 55 in 1990 are based on the compilation of historical estimates of the components

Table 4. Alternative Uncertainty Intervals for the Demographic Analysis Estimates of Percent Net Undercount by Race, Sex, and Age: 1990

Race, Sex, Age (years)	Percent undercount		95% Intervals			99% Intervals		
	Observed	Mean	Lower	Upper	Length	Lower	Upper	Length
Black male								
0–9	8.07	8.59	5.96	11.22	5.26	4.34	12.84	8.51
10–19	1.95	2.51	.36	4.65	4.30	−.88	5.89	6.77
20–29	9.09	10.08	8.35	11.82	3.47	7.41	12.76	5.35
30–44	12.50	13.55	11.63	15.47	3.83	10.53	16.57	6.03
45–64	11.87	13.44	9.15	17.74	8.59	6.32	20.56	14.24
65+	3.00	2.34	−1.44	6.13	7.56	−3.88	8.57	12.44
Total	8.47	9.31	7.18	11.44	4.25	5.92	12.70	6.78
Black female								
0–9	7.75	8.21	5.63	10.79	5.16	4.00	12.41	8.41
10–19	2.13	2.62	.56	4.68	4.12	−.66	5.89	6.54
20–29	3.47	4.39	2.68	6.11	3.43	1.73	7.06	5.33
30–44	2.55	3.63	1.60	5.66	4.06	.41	6.85	6.44
45–64	.61	2.29	−2.07	6.64	8.72	−4.94	9.51	14.46
65+	−.95	1.58	−1.60	4.76	6.36	−3.64	6.80	10.44
Total	2.97	4.03	1.94	6.12	4.18	.69	7.37	6.69
Non-Black male								
0–9	2.63	3.19	2.34	4.03	1.69	1.79	4.59	2.80
10–19	−.89	−.16	−1.11	.79	1.90	−1.74	1.42	3.16
20–29	1.70	2.68	1.47	3.90	2.42	.66	4.71	4.05
30–44	2.89	3.85	2.70	5.00	2.30	1.92	5.78	3.85
45–64	2.73	2.93	.87	4.99	4.12	−.52	6.39	6.92
65+	1.42	.84	−1.14	2.83	3.97	−2.49	4.17	6.66
Total	1.94	2.51	1.49	3.52	2.04	.81	4.21	3.40
Non-Black female								
0–9	2.76	3.33	2.49	4.16	1.67	1.94	4.71	2.77
10–19	−.53	.17	−.73	1.07	1.80	−1.32	1.66	2.99
20–29	.63	1.47	.42	2.52	2.10	−.28	3.22	3.50
30–44	.22	1.14	−.09	2.36	2.45	−.91	3.19	4.10
45–64	.44	.70	−1.45	2.84	4.29	−2.90	4.29	7.20
65+	.40	1.24	−.43	2.92	3.35	−1.57	4.05	5.62
Total	.61	1.30	.29	2.31	2.03	−.39	2.99	3.39
Total population								
0–9	3.53	4.08	3.08	5.08	2.00	2.40	5.76	3.35
10–19	−.28	.40	−.55	1.35	1.90	−1.19	1.99	3.18
20–29	1.90	2.81	1.65	3.97	2.33	.86	4.76	3.90
30–44	2.30	3.25	2.14	4.37	2.23	1.38	5.12	3.74
45–64	2.02	2.40	.67	4.13	3.45	−.50	5.30	5.79
65+	.79	1.14	−.68	2.97	3.66	−1.92	4.21	6.14
Total	1.83	2.49	1.63	3.36	1.73	1.04	3.95	2.90

NOTE The 95% uncertainty intervals represent an error model with a 95% uncertainty interval and multiplier limits defined as 99.9% certain. The 99% uncertainty intervals represent an error model with a broader 99% uncertainty interval and multiplier limits defined as 99% certain.

data. Thus, although the undercount rates for Black males and Black females for 1990 are widely different (8.47 and 2.97), the corresponding lengths of the error intervals are about the same (4.25 and 4.18).

The means of the percent net undercount in Table 4 clearly indicate that the demographic net undercount estimates are biased in that they may underestimate the "true" net undercount (compare the estimates in column 1 and column 2). In fact, for the younger age groups of Non-Blacks, the undercount estimates fall close to the lower bounds of the 95% error intervals. For example, the demographic "point" estimate of .63% for Non-Black females age 20–29 is near the lower bound estimate of .42, the estimated mean of 1.47% being more than double the point estimate.

4.6 Limitations of the Demographic Uncertainty Estimates

The systematic and detailed evaluation of the quality of the demographic coverage estimates reported here represents an evaluation program *new* for the 1990 census. The assessments conducted in the 11 evaluation projects are subject to change and improvement over time just as the basic demographic estimates have been. But we feel that the models, assumptions, and analysis of the available information for the evaluation projects provide a reasonable assessment of the overall uncertainty in the demographic estimates of population and coverage for the 1990 census.

5. CONCLUSIONS

The technique of demographic analysis is a powerful tool for measuring net undercount in a census. The 1990 demographic analysis program provided not only the completeness of census coverage based on demographic analysis but also an assessment of the quality of these coverage estimates.

The estimates of net undercount for particular race, sex, or age groups based on demographic analysis may be subject to considerable uncertainty for measuring the exact *levels*. But they are subject to *less* variability in terms of measuring *differences* in coverage according to age, sex, and race and

ESCAP MEETING NO. 44 - 02/22/01

MINUTES

Minutes of the Executive Steering Committee on Accuracy and Coverage Evaluation (A.C.E.) Policy (ESCAP) Meeting # 44

February 22, 2001

Prepared by: Sarah Brady

The forty-fourth meeting of the Executive Steering Committee on Accuracy and Coverage Evaluation Policy was held on February 22, 2001 at 10:30 and 2. The agenda for the 10:30 meeting was to discuss concerns and unresolved issues. The agenda for the 2:00 meeting was to discuss results from demographic analysis with a doubling of undocumented immigration.

Committee Attendees:

Nancy Potok
Paula Schneider
Cynthia Clark
Nancy Gordon
John Thompson
Jay Waite
Bob Fay
Howard Hogan
Ruth Ann Killion
John Long
Carol Van Horn

Deputy Director/Acting Director:
William Barron

Other Attendees:

Marvin Raines
William Bell
Gregg Robinson (PM only)
Kathleen Styles (AM only)
Maria Urrutia
Sarah Brady

I. Concerns and Unresolved Issues

The Committee discussed concerns individual members had that were essential to his/her decision about whether adjustment would improve the accuracy of Census 2000. The major concerns cited were as follows:

- Explaining the difference between DA and the A.C.E.
- The synthetic assumption.
- The ability to interpret results from the loss functions in terms of the degree of improvement provided by the A.C.E.-Bill Bell and Howard Hogan will do work on deriving some measures.
- The impact of A.C.E. on small areas and groups.
- The construct of the race/Hispanic origin domains- Need to carefully explain how they were constructed to the data users.
- Targeted Extended Search (TES) and its related balancing issues.
- Concerns that the 1990 Post Enumeration Survey (PES) potentially underestimated the undercount because correlation bias was larger in 1990 than measured.
- Concerns about the models for correlation bias since it has a significant influence on the analysis, therefore, requiring more sensitivity analysis.
- The time available to the Committee to complete their review.
- Concerns were raised to not let perfection be the overriding goal. Rather, the Committee should determine if an improvement can be made to the accuracy of the census with the A.C.E. results.

II. Demographic Analysis

Gregg Robinson presented a revised version of the demographic analysis (DA) results from the DSSD Census 2000 Procedures and Operations Memorandum Series B-4, which was presented to the ESCAP on February 7, 2001. The revised DA report is attached. The revised DA incorporated a doubling of the estimated undocumented immigration population. This revision caused DA to demonstrate a net undercount rather than a net overcount as previously presented. However, the revised DA estimate still did not agree with the A.C.E. estimate from the Dual System Estimation.

The Committee also discussed a paper Gregg had prepared in 1993 in the Journal of the American Statistical Association. The paper discussed that there was a wide range of uncertainty to the DA estimates in 1990. John Long and Howard Hogan will research this further by examining the possibility that the 1990 census, PES, and 1990 DA missed a significant portion of the population due to correlation bias being much larger in 1990 than believed. As a result, the undercounts in 1990 were an underestimation.

III. Next Meeting

The agenda for the next meeting, scheduled for February 23, 2001, is to discuss results from the sensitivity analysis of the loss functions.

ESCAP MEETING NO. 45 - 02/23/01

AGENDA

There was no agenda developed or used for the February 23, 2001 meeting.

ESCAP MEETING NO. 45 - 02/23/01

HANDOUTS

Materials attached to these minutes were draft and preliminary material to inform the ESCAP Committee. The data and analysis contained in these documents are subject to revision and are not final. These materials report the results of research and analysis undertaken by Census Bureau staff. They have undergone a more limited review than official Census Bureau publications. Research results and conclusions expressed are those of the authors and do not necessarily indicate concurrence by the Census Bureau.

Table 2. Sources of Data for Estimation of Components of Error

Error Components	Measurement in 1990	Measurement in 2000
P-sample matching error	1990 Matching Error Study	1990 Matching Error Study with adjustments for 2000
P-sample data collection error	1990 Evaluation Followup	1990 Evaluation Followup with adjustments for 2000
P-sample fabrication	1990 Evaluation Followup	1990 Evaluation Followup with adjustments for 2000
E-sample data collection error	1990 Evaluation Followup	1990 Evaluation Followup with adjustments for 2000
E-sample processing error	1990 Matching Error Study	1990 Matching Error Study with adjustments for 2000
Correlation bias	1990 Demographic Analysis	2000 Demographic Analysis
Ratio estimator bias	1990 PES	2000 A.C.E.
Sampling error	1990 PES	2000 A.C.E.
Imputation error	1990 Reasonable Alternatives Imputation Study	1990 Reasonable Alternatives with adjustments for 2000
Excluded Census Data Error	1990 Excluded Data Study	Not available
Contamination of P sample by enumeration or vice versa	Shown to be negligible	Not available in time for analysis for decision
Misclassification error of records into poststrata from inconsistent reporting	Not measured	Not available in time for analysis for decision
Synthetic error	Artificial population analysis and not integrated in total error model	Under development but will not be integrated in total error model

16 Evaluation Poststrata

		No. in MVF P-sample (1990)	PS Groups (2000)
1. Non-minority/owner/large and Medium MSA MO-MB NE/MW	high RR	4,960	1,2,9,10
2. Non-minority/owner/large and Medium MSA MO-MB S/W	high RR	7,702	3,4,11,12
3. Non-minority/owner/large and Medium MSA MO-MB NE/MW	low RR	3,031	5,6,13,14
4. Non-minority/owner/large and Medium MSA MO-MB S/W	low RR	2,936	7,8,15,16
5. Non-minority/owner/Small MSA and Non-MSA MO-MB	high RR	5,560	17-20
6. Non-minority/owner/ Small MSA and Non-MSA MO-MB	low RR	2,095	21-24
7. Non-minority/Owner/All Other TEAs		7,355	25-32
8. Non-minority/ Non-Owner/Large or Medium MSA MO-MB	high RR	4,963	33, 35
9.Non-minority/ Non-Owner/Large or Medium MSA MO-MB	low RR	3,197	34, 36
10. Non-minority/non-owner/Small MSA & Non-MSA MO-MB All other TEA		5,291	37-40
11. Minority/owner/large and Medium MSA MO-MB	high RR	8,841	41, 49, 57,59
12. Minority/owner/large and Medium MSA MO-MB	low RR	5,628	42, 50
13. Minority/Owner/All Other TEAs		3,877	43, 44, 51, 52
14. Minority/ Non-Owner/Large or Medium MSA MO-MB	high RR	10,809	45, 53, 58,60
15. Minority/ Non-Owner/Large or Medium MSA MO-MB	low RR	6,421	46, 54
16. Minority/Non-Owner/All Other TEAs		3,797	47, 48, 55, 56, 61-64
Total		86,463	

Table 20.	Total error for net undercount rate assuming no correlation bias						
Ev post	prod uc	corr uc	bias(uc)	se(bias)	se(pr uc)	total se	95% conf interval
US	1.1788	0.464	0.7148	0.086	0.1349	0.16	(0.1439, 0.7840)
1	0.2695	-0.1217	0.3912	0.2181	0.224	0.3127	(-0.7470, 0.5036)
2	0.0947	-0.2516	0.3463	0.1802	0.255	0.3123	(-0.8761, 0.3729)
3	-2.8191	-5.2887	2.4696	0.4999	0.6572	0.8257	(-6.9401, -3.6373)
4	1.284	1.9862	-0.7022	0.6324	0.9813	1.1674	(-0.3486, 4.3209)
5	0.2127	-0.207	0.4197	0.1047	0.2792	0.2982	(-0.8034, 0.3894)
6	2.3302	1.8476	0.4826	0.2876	0.793	0.8436	(0.1605, 3.5347)
7	0.4232	-0.853	1.2762	0.2266	0.3562	0.4222	(-1.6973, -0.0087)
8	1.129	1.0745	0.0545	0.1754	0.4918	0.5221	(0.0303, 2.1187)
9	1.8404	1.2102	0.6302	0.3538	0.7921	0.8675	(-0.5248, 2.9453)
10	2.5867	1.5337	1.053	0.54	0.4426	0.6982	(0.1373, 2.9302)
11	1.3307	0.3131	1.0177	0.2522	0.3897	0.4642	(-0.6153, 1.2414)
12	-0.6778	-1.7953	1.1176	0.4734	0.8642	0.9853	(-3.7660, 0.1753)
13	0.7719	-0.3231	1.095	0.4806	0.6944	0.8445	(-2.0120, 1.3659)
14	3.5018	3.1517	0.3502	0.386	0.4592	0.5999	(1.9519, 4.3515)
15	4.214	2.8633	1.3507	0.4191	0.8036	0.9063	(1.0506, 4.6760)
16	3.9699	1.7715	2.1984	0.4931	0.7404	0.8895	(-0.0075, 3.5505)

Table 22	Total error of net undercount rate assuming no correlation bias for 18-29 Nonblack males						
Ev post	prod uc	corr uc	bias(uc)	se(bias)	se(prod uc)	total se	95% conf interval
US	1.1788	0.8567	0.3221	0.0857	0.1349	0.1598	(0.5370, 1.1763)
1	0.2695	0.0413	0.2282	0.2176	0.224	0.3123	(-0.5834, 0.6660)
2	0.0947	-0.0766	0.1714	0.18	0.255	0.3121	(-0.7009, 0.5477)
3	-2.8191	-5.1012	2.2821	0.499	0.6572	0.8252	(-6.7516, -3.4508)
4	1.284	2.1584	-0.8745	0.6308	0.9813	1.1665	(-0.1746, 4.4915)
5	0.2127	-0.0282	0.2409	0.1046	0.2792	0.2982	(-0.6245, 0.5681)
6	2.3302	2.0222	0.308	0.2867	0.793	0.8432	(0.3357, 3.7086)
7	0.4232	-0.6727	1.0958	0.226	0.3562	0.4219	(-1.5164, 0.1710)
8	1.129	1.1742	-0.0452	0.175	0.4918	0.522	(0.1302, 2.2182)
9	1.8404	1.3175	0.523	0.3535	0.7921	0.8674	(-0.4173, 3.0522)
10	2.5867	1.6215	0.9652	0.5396	0.4426	0.6979	(0.2257, 3.0173)
11	1.3307	1.1906	0.1402	0.2497	0.3897	0.4628	(0.2649, 2.1163)
12	-0.6778	-0.4871	-0.1907	0.4674	0.8642	0.9824	(-2.4519, 1.4778)
13	0.7719	0.9453	-0.1734	0.4792	0.6944	0.8437	(-0.7421, 2.6327)
14	3.5018	3.9754	-0.4736	0.3819	0.4592	0.5973	(2.7808, 5.1699)
15	4.214	3.843	0.371	0.4147	0.8036	0.9043	(2.0345, 5.6516)
16	3.9699	2.6335	1.3364	0.4888	0.7404	0.8871	(0.8592, 4.4078)

Table 23	Total error for net undercount rate						
	assuming correlation bias of 2% overcount for 18-29 NB males						
Ev post	prod uc	corr uc	bias(uc)	se(bias)	se(prod uc)	total se	95% conf interval
US	1.1788	0.721	0.4578	0.0857	0.1349	0.1598	(0.4013, 1.0407)
1	0.2695	-0.0548	0.3243	0.2179	0.224	0.3125	(-0.6798, 0.5702)
2	0.0947	-0.1638	0.2585	0.1801	0.255	0.3122	(-0.7882, 0.4606)
3	-2.8191	-5.2214	2.4023	0.4996	0.6572	0.8256	(-6.8725, -3.5703)
4	1.284	2.0528	-0.7689	0.6313	0.9813	1.1668	(0.2808, 4.3865)
5	0.2127	-0.1252	0.3379	0.1047	0.2792	0.2982	(-0.7215, 0.4712)
6	2.3302	1.9149	0.4153	0.2869	0.793	0.8433	(0.2283, 3.6015)
7	0.4232	-0.7775	1.2007	0.2262	0.3562	0.422	(-1.6214, 0.0664)
8	1.129	0.929	0.2	0.1757	0.4918	0.5222	(-0.1155, 1.9734)
9	1.8404	1.0214	0.819	0.3555	0.7921	0.8682	(-0.7150, 2.7578)
10	2.5867	1.3541	1.2326	0.541	0.4426	0.699	(-0.0438, 2.7520)
11	1.3307	1.0932	0.2376	0.2494	0.3897	0.4627	(0.1679, 2.0185)
12	-0.6778	-0.5841	-0.0937	0.4667	0.8642	0.9821	(-2.5483, 1.3802)
13	0.7719	0.8665	-0.0946	0.4775	0.6944	0.8427	(-0.8190, 2.5520)
14	3.5018	3.806	-0.3042	0.3821	0.4592	0.5974	(2.6112, 5.0008)
15	4.214	3.6797	0.5343	0.4154	0.8036	0.9046	(1.8705, 5.4889)
16	3.9699	2.4846	1.4853	0.4893	0.7404	0.8874	(0.7097, 4.2594)

Figure 2

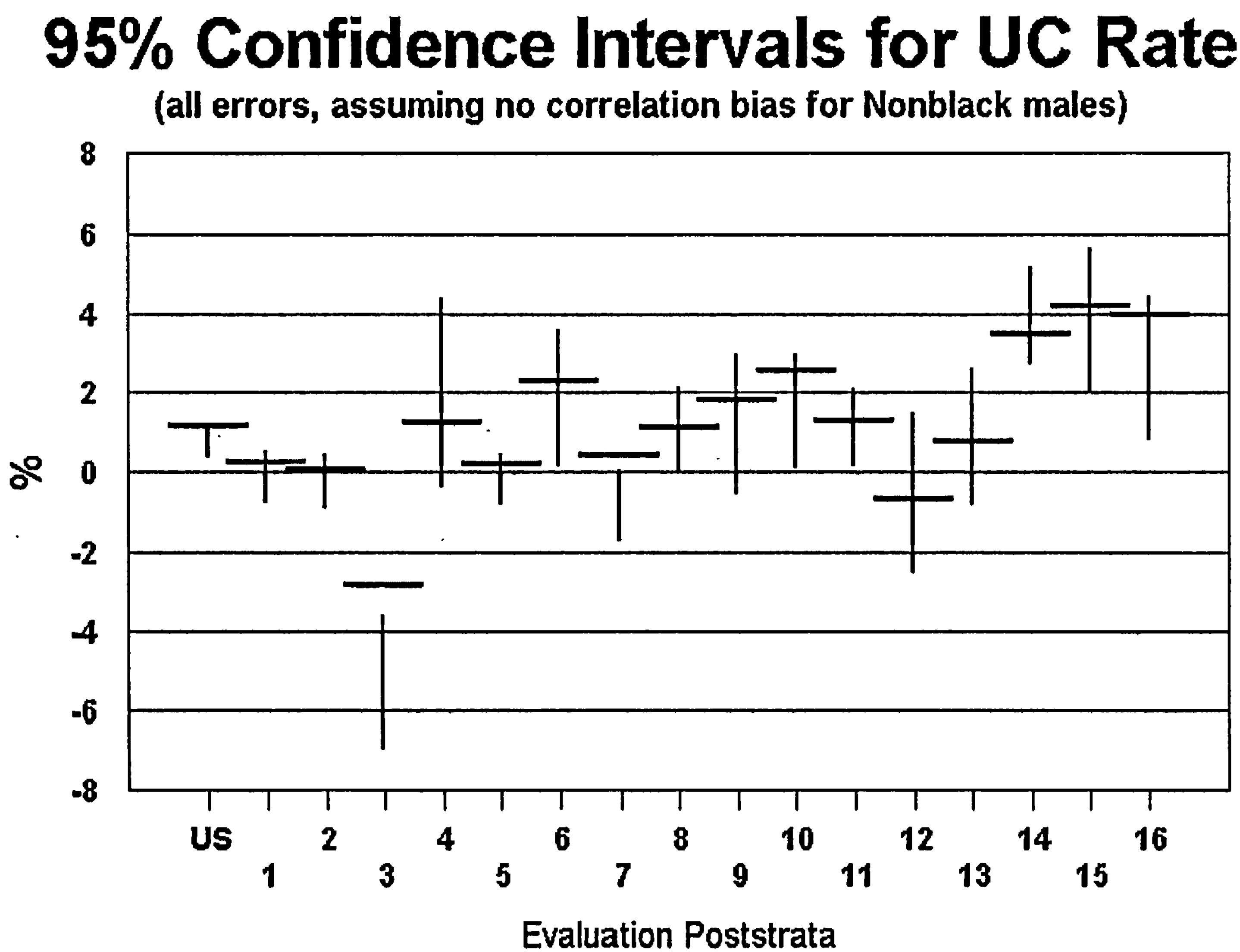
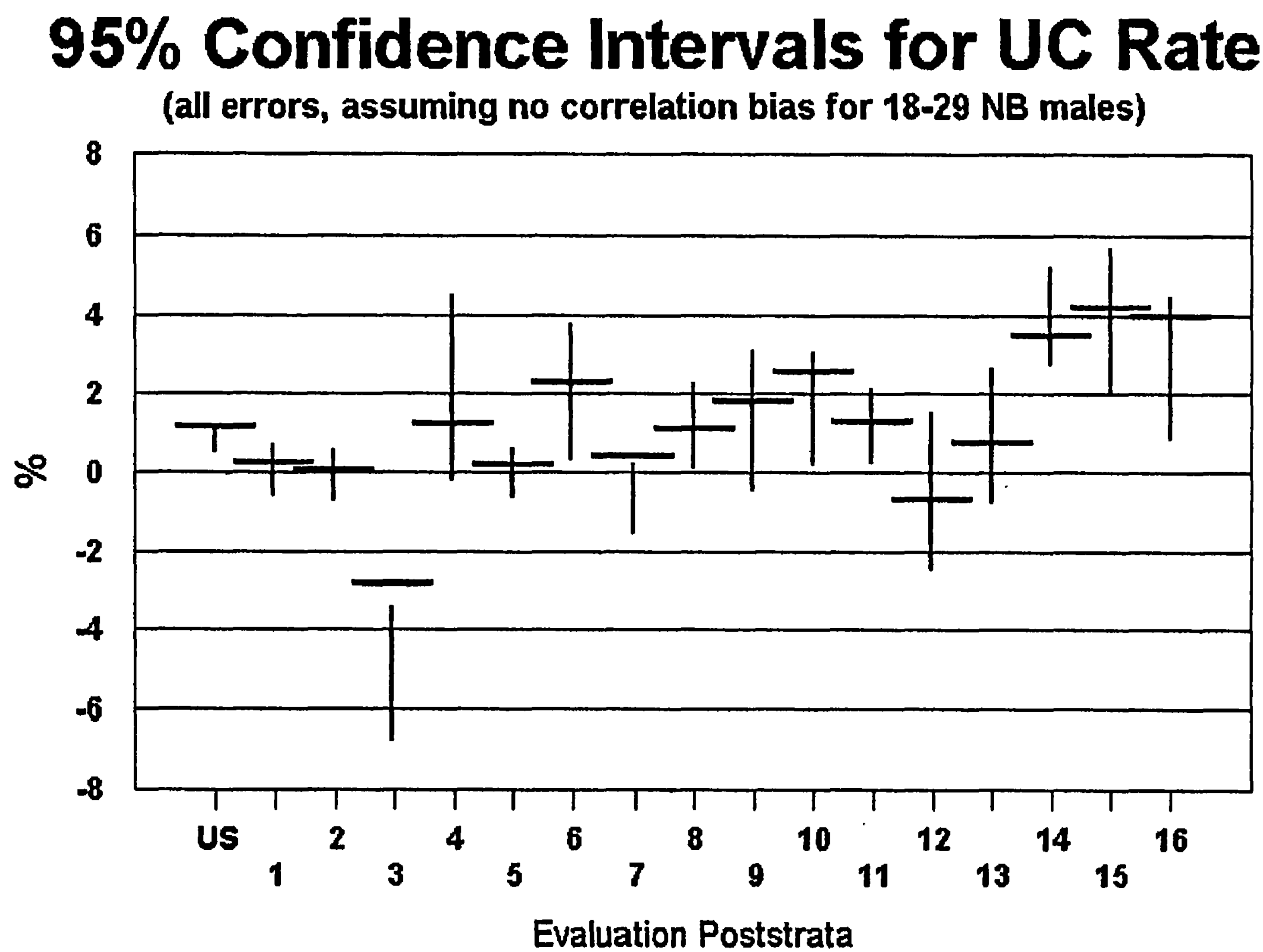

95% Confidence Intervals for UC Rate
(all errors, assuming no correlation bias for Nonblack males)
%
8
6
4
2
0
-2
-4
-6
-8
US
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
Evaluation Poststrata

Figure 3



95% Confidence Intervals for UC Rate
(all errors, assuming no correlation bias for 18-29 NB males)
%
8
6
4
2
0
-2
-4
-6
-8
US
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
Evaluation Poststrata

Moments of Error Components for
Evaluation Poststratum 01
Non-min/owner/Large or Medium
MSA - High - NE/MW

Nce = 34,282,550

Direct DSE = 35,646,814

Np = 35,984,939

U = 0.271

Ncp = 34,607,734

Census = 35,550,177

Error Source	Pos Gross Error	Neg Gross Error	Net Error	B(U)
Matching Error				
Mm	79,467	160,433	(80,966)	0.29
	(46,175)	(49,853)	(58,746)	
Npm	71,576	49,913	21663	
	(27,714)	(25,519)	(10,508)	
P Sample Collection				
Npa	297,615	38,881	258,734	0.28
	(232,050)	(16,135)	(231,518)	
Ma	242,628	89,648	152,980	
	(229,502)	(39,140)	(234,747)	
P Sample Fabrication				
Npf		0	0	0.00
		0	0	
Mf		0	0	
		0	0	
E Sample Error				
Co	118,562	104,287	14,275	0.04
	(67,514)	(49,188)	(83,466)	
Cc	50,031	128,833	(78,803)	-0.23
	(19,471)	(67,035)	(70,132)	
Model Bias (Tau)				
Imputation Error				
Ratio Estimator Bias			4,608	0.01
			461	
Net Sampling Error			0	
			(80,076)	

Moments of Error Components for
Evaluation Poststratum 02

Non-min/owner/Large or Medium
MSA - High - S/W

Nce = 29,785,508 Direct DSE = 31,284,669

Np = 32,448,004 U = 0.102

Ncp = 30,893,095 Census = 31,252,841

Error Source	Pos Gross Error	Neg Gross Error	Net Error	B(U)
Matching Error				
Mm	65,321	95,909	-30588	0.01
	(32,990)	(33,361)	(46,830)	
Npm	35,732	64,360	-28628	
	(15,551)	(33,111)	(36,549)	
P Sample Collection				
Npa	59,015	36,486	22,529	0.33
	(29,549)	(17,011)	(18,124)	
Ma	41,316	120,654	-79339	
	(41,316)	(66,891)	(38,580)	
P Sample Fabrication				
Npf		5,168	5,168	0.04
		(5,168)	(5,168)	
Mf		17,377	17,377	
		(17,377)	(17,377)	
E Sample Error				
Co	55,759	36,447	19,312	0.06
	(36,038)	(13,756)	(38,438)	
Cc	34,409	65,455	-31,046	-0.10
	(25,522)	(31,415)	(40,787)	
Model Bias (Tau)				
Imputation Error				
Ratio Estimator Bias			553	0.01
			55	
Net Sampling Error			0	
			(79,854)	

Moments of Error Components for
Evaluation Poststratum 03
Non-min/owner/Large or Medium
MSA - Low - NE/MW

Nce = 4,893,801 | Direct DSE = 5,217,719

Np = 4,811,175 | U = -2.941

Ncp = 4,512,495 | Census = 5,371,168

Error Source	Pos Gross Error	Neg Gross Error	Net Error	B(U)
Matching Error				
Mm	6,283	14,788	-8,504	0.31
	(4,191)	(6,554)	(5,396)	
Npm	7,770	2,366	5,404	
	(5,290)	(1,579)	(5,545)	
P Sample Collection				
Npa	9,723	8,197	1,526	0.75
	(5,992)	(5,605)	(7,203)	
Ma	379	31,983	-31,604	
	(379)	(19,963)	(19,967)	
P Sample Fabrication				
Npf		0	0	0.00
		(0)	(0)	
Mf		0	0	
		(0)	(0)	
E Sample Error				
Co	71,698	13,777	57,921	1.23
	(49,900)	(6,220)	(50,269)	
Cc	29,353	23,113	6,239	0.13
	(15,591)	(13,733)	(17,748)	
Model Bias (Tau)				
Imputation Error				
Ratio Estimator Bias			120	0.00
			(12)	
Net Sampling Error			0	
			(33,391)	

Moments of Error Components for

Evaluation Poststratum 04
Non-min/owner/Large or Medium
MSA -Low - S/W

Nce = 7,716,712 Direct DSE = 8,358,892

Np = 7,173,516 U = 1.275

Ncp = 6,622,403 Census = 8,252,306

Error Source	Pos Gross Error	Neg Gross Error	Net Error	B(U)
Matching Error				
Mm	28,951	19,464	9,487	-0.08
	(13,219)	(9,074)	(16,333)	
Npm	23,231	18,491	4,741	
	(13,969)	(9,105)	(13,041)	
P Sample Collection				
Npa	123,199	122,715	484	0.06
	(111,652)	(107,867)	(10,659)	
Ma	106,748	110,215	-3,468	
	(105,819)	(109,025)	(51,235	
P Sample Fabrication				
Npf		0	0	0.00
		(0)	(0)	
Mf		0	0	
		(0)	(0)	
E Sample Error				
Co	45,454	47,054	-1,600	-0.02
	(19,432)	(25,323)	(32,585)	
Cc	2,216	53,934	-51,718	-0.66
	(1,989)	(39,686)	(39,736)	
Model Bias (Tau)				
Imputation Error				
Ratio Estimator Bias			1,113	0.01
			(111)	
Net Sampling Error			0	
			(83,149)	

Moments of Error Components for
Evaluation Poststratum 05

Non-min/owner/Small and Non-MSA - High

Nce = 24,649,632 | Direct DSE = 25,751,566

Np = 25,377,448 | U = 0.209

Ncp = 24,291,523 | Census = 25,697,696

Error Source	Pos Gross Error	Neg Gross Error	Net Error	B(U)
Matching Error				
Mm	31,555	63,078	-31,523	0.17
	(13,256)	(21,561)	(24,222)	
Npm	29,184	18,333	10,851	
	(11,944)	(8,293)	(14,661	
P Sample Collection				
Npa	29,131	20,600	8,531	0.08
	(13,518)	(13,855)	(12,258)	
Ma	6,513	16,627	-10,114	
	(6,513)	(9,888)	(11,839)	
P Sample Fabrication				
Npf		0	0	0.00
		(0)	(0)	
Mf		0	0	
		(0)	(0)	
E Sample Error				
Co	94,940	33,672	61,269	0.25
	(35,208)	(18,906)	(37,892)	
Cc	21,980	43,677	-21,697	-0.09
	(11,606)	(18,101)	(21,518)	
Model Bias (Tau)				
Imputation Error				
Ratio Estimator Bias			11,291	0.04
			(1,129)	
Net Sampling Error			0	
			(72,062)	

Moments of Error Components for
Evaluation Poststratum 06
Non-min/owner/Small and Non-

MSA - Low

Nce = 5,817,573 Direct DSE = 6,338,959

Np = 6,441,327 U = 2.203

Ncp = 5,911,522 Census = 6,199,286

Error Source	Pos Gross Error	Neg Gross Error	Net Error	B(U)
Matching Error				
Mm	16,208	21,084	-4,876	0.06
	(9,084)	(12,779)	(14,539)	
Npm	14,590	16,007	-1,417	
	(9,597)	(8,857)	(7,006)	
P Sample Collection				
Npa	-24,405	17,576	6,828	0.36
	(16,086)	(10,715)	(6,828)	
Ma	8,856	24,542	-15,686	
	(6,418)	(13,039)	(10,292)	
P Sample Fabrication				
Npf		0	0	0.00
		(0)	(0)	
Mf		0	0	
		(0)	(0)	
E Sample Error				
Co	61,037	45,235	15,802	0.27
	(34,009)	(22,425)	(17,298)	
Cc	16,105	33,286	-17,180	-0.29
	(15,250)	(21,584)	(9,343)	
Model Bias (Tau)				
Imputation Error				
Ratio Estimator Bias			415	0.01
			(42)	
Net Sampling Error			0	
			(51,526)	

Moments of Error Components for
Evaluation Poststratum 07
Non-min/owner/All Other TEAs

Nce = 32,195,096 | Direct DSE = 34,773,055

Np = 32,656,527 | U = 0.401

Ncp = 30,235,481 | Census = 34,633,612

Error Source	Pos Gross Error	Neg Gross Error	Net Error	B(U)
Matching Error				
Mm	85,548	168,896	-83,349	0.23
	(40,951)	(46,856)	(38,564)	
Npm	83,393	96,893	-13,501	
	(31,865)	(43,863)	(21,688)	
P Sample Collection				
Npa	160,135	202,065	-41,930	0.79
	(44,847)	(82,204)	(73,861)	
Ma	42,081	320,702	-278,621	
	(15,982)	(132,414)	(131,817)	
P Sample Fabrication				
Npf		0	0	0.00
		(0)	(0)	
Mf		0	0	
		(0)	(0)	
E Sample Error				
Co	269,788	114,722	155,067	0.48
	(125,846)	(42,075)	(109,567)	
Cc	24,602	100,381	-75,778	-0.23
	(15,546)	(39,392)	(34,993)	
Model Bias (Tau)				
Imputation Error				
Ratio Estimator Bias			13,366	0.04
			(1,337)	
Net Sampling Error			0	
			(124,413)	

Moments of Error Components for Evaluation Poststratum 08 Non-min/non-owner/Large or Medium MSA - High

Nce = 18,112,506 | Direct DSE =

20,213,083

Np = 19,175,297 U = 1.097

Ncp = 17,182,568 Census = 19,991,324

Error Source	Pos Gross Error	Neg Gross Error	Net Error	B(U)
Matching Error				
Mm	34,026	26,913	7,113	0.01
	(14,091)	(15,149)	(17,050)	
Npm	29,384	20,120	9,264	
	(13,256)	(10,935)	(10,535)	
P Sample Collection				
Npa	63,198	93,825	-30,627	0.38
	(23,359)	(40,768)	(29,128)	
Ma	24,341	117,647	-93,305	
	(13,343)	(69,209)	(57,457)	
P Sample Fabrication				
Npf		0	0	0.00
		(0)	(0)	
Mf		0	0	
		(0)	(0)	
E Sample Error				
Co	42,893	44,076	-1,183	-0.01
	(16,008)	(24,311)	(28,123)	
Cc	47,640	105,936	-58,296	-0.32
	(19,356)	(41,402)	(46,174)	
Model Bias (Tau)				
Imputation Error				
Ratio Estimator Bias			-1,264	-0.01
			126	
Net Sampling Error			0	
			(100,570)	

Moments of Error Components for
Evaluation Poststratum 09
Non-min/non-owner/Large or
Medium MSA - Low

Nce = 6,023,062 Direct DSE =
7,035,171

Matching Error				
Mm	46,200	78,906	-32,706	0.23
	(20,419)	(32,231)	(37,871)	
Npm	38,015	56,210	-18,195	
	(30,391)	(32,901)	(15,061)	
P Sample Collection				
Npa	42,315	14,861	27,454	1.55
	(14,485)	(13,068)	(19,479)	
Ma	11,339	102,076	-90,737	
	(7,899)	(46,791)	(47,398)	
P Sample Fabrication				
Npf		9,366	-9,366	0.00
		(9,366)	(9,366)	
Mf		8,070	-8,070	
		(8,070)	(8,070)	
E Sample Error				
Co	74,094	47,069	27,024	0.32
	(27,839)	(24,685)	(37,285)	
Cc	51,700	49,975	1,726	0.02
	(24,020)	(32,581)	(36,183)	
Model Bias (Tau)				
Imputation Error				
Ratio Estimator Bias			1,235	0.01
			(124)	
Net Sampling Error			0	
			(74,985)	

Np = 6,468,268 U = 1.799

Ncp = 5,537,716 Census = 6,908,574

Error Source	Pos Gross Error	Neg Gross Error	Net Error	B(U)
Matching Error				
Mm	19,705	41,141	-21,436	0.50
	(7,604)	(14,417)	(15,543)	
Npm	24,135	16,092	8,044	
	(12,088)	(6,618)	(13,428)	
P Sample Collection				
Npa	19,360	39,218	-19,858	0.16
	(9,346)	(33,993)	(34,962)	
Ma	8,439	34,344	-25,905	
	(5,976)	(32,107)	(32,582)	
P Sample Fabrication				
Npf		0	0	0.00
		(0)	(0)	
Mf		0	0	
		(0)	(0)	
E Sample Error				
Co	62,054	33,753	28,300	0.46
	(22,099)	(12,581)	(21,750)	
Cc	18,456	49,762	-31,306	-0.51
	(8,243)	(20,936)	(22,541)	
Model Bias (Tau)				
Imputation Error				
Ratio Estimator Bias			322	0.00
			(32)	
Net Sampling Error			0	
			(56,793)	

Moments of Error Components for
Evaluation Poststratum 10
Non-min/non-owner/Small and
Non-MSA, All Other TEAs

Nce = 17,212,267 Direct DSE = 19,551,600

Np = 18,265,774

U = 2.479

Ncp = 16,080,289

Census = 19,067,004

Error Source	Pos Gross Error	Neg Gross Error	Net Error	B(U)
Matching Error				
Mm	51,055	206,208	-155,153	1.07
	(16,751)	(80,344)	(79,702)	
Npm	67,768	44,465	23,304	
	(27,222)	(17,623)	(23,523)	
P Sample Collection				
Npa	84,539	98,049	-13,510	0.62
	(25,718)	(33,980)	(27,515)	
Ma	24,531	138,758	-114,227	
	(11,665)	(42,276)	(39,571)	
P Sample Fabrication				
Npf		0	0	0.00
		(0)	(0)	
Mf		0	0	
		(0)	(0)	
E Sample Error				
Co	196,274	93,909	102,366	0.58
	(79,354)	(32,660)	(85,676)	
Cc	136,021	351,845	-215,824	-1.21
	(50,836)	(190,221)	(196,523)	
Model Bias (Tau)				
Imputation Error				
Ratio Estimator Bias			4,381	0.02
			(438)	
Net Sampling Error			0	
			(88,933)	

Moments of Error Components for Evaluation Poststratum 11 Minority/owner/Large or Medium MSA - High

Nce = 22,815,631

Direct DSE = 24,896,228

Np = 23,316,868

U = 1.284

Ncp = 21,368,260 Census = 24,576,535

Error Source	Pos Gross Error	Neg Gross Error	Net Error	B(U)
Matching Error				
Mm	32,919	69,315	-36,396	0.10
	(10,981)	(23,217)	(25,823)	
Npm	22,469	37,631	-15,162	
	(10,278)	(10,799)	(14,812)	
P Sample Collection				
Npa	228,619	34,430	194,189	0.67
	(100,750)	(15,467)	(99,953)	
Ma	103,727	70,384	33,343	
	(92,085)	(30,703)	(95,865)	
P Sample Fabrication				
Npf		292	-292	0.00
		(292)	(292)	
Mf		269	-269	
		(269)	(269	
E Sample Error				
Co	74,568	16,690	57,878	0.25
	(38,564)	(6,536)	(39,074)	
Cc	39,425	46,790	-7,365	-0.03
	(11,436)	(21,650)	(23,803)	
Model Bias (Tau)				
Imputation Error				
Ratio Estimator Bias			8,683	0.03
			(868)	
Net Sampling Error			0	
			(98,342)	

Moments of Error Components for
Evaluation Poststratum 12
Minority/owner/Large or Medium
MSA - Low

Nce = 4,620,389 Direct DSE = 5,285,962

Np = 4,532,239 U = -0.765

Ncp = 3,961,569 Census = 5,326,380

Error Source	Pos Gross Error	Neg Gross Error	Net Error	B(U)
Matching Error				
Mm	9,165	22,438	-13,273	0.61
	(6,192)	(9,053)	(10,999)	
Npm	19,938	7,767	12,171	
	(9,556)	(6,086)	(11,328)	
P Sample Collection				
Npa	9,666	51,283	-41,617	0.52
	(3,946)	(40,001)	(40,202)	
Ma	406	57,469	-57,063	
	(406)	(43,042)	(43,044)	
P Sample Fabrication				
Npf		0	0	0.00
		(0)	(0)	
		0	0	
		(0)	(0)	
E Sample Error				
Co	16,157	14,033	2,124	0.05
	(4,890)	(3,901)	(4,955)	
Cc	14,196	19,248	-5,053	-0.11
	(4,226)	(6,144)	(6,927)	
Model Bias (Tau)				
Imputation Error				
Ratio Estimator Bias			296	0.01
			(30)	
Net Sampling Error			0	
			(45,411)	

Moments of Error Components for
Evaluation Poststratum 13
Minority/owner/All Other TEAs

Nce = 8,859,679 Direct DSE = 9,841,047

Np = 8,697,210 U = 0.651

Ncp = 7,829,907 Census = 9,776,940

Error Source	Pos Gross Error	Neg Gross Error	Net Error	B(U)
Matching Error				
Mm	18,377	13,996	4,382	-0.19
	(8,636)	(7,431)	(11,354)	
Npm	10,642	22,324	-11,681	
	(5,917)	(11,875)	(13,130	
P Sample Collection				
Npa	183,215	130,607	52,607	0.43
	(133,248)	(130,607)	(27,946)	
Ma	22,904	9,264	13,640	
	(22,904)	(6,598)	(23,887)	
P Sample Fabrication				
Npf		0	0	0.00
		(0)	(0)	
Mf		0	0	
		(0)	(0)	
E Sample Error				
Co	31,697	21,524	10,173	0.11
	(13,204)	(10,925)	(17,258)	
Cc	78,809	14,080	64,728	0.73
	(31,658)	(7,584)	(33,160)	
Model Bias (Tau)				
Imputation Error				
Ratio Estimator Bias			1,140	0.01
			(114)	
Net Sampling Error			0	
			(68,949)	

Moments of Error Components for
Evaluation Poststratum 14
Minority/non-owner/Large or
Medium MSA - High

Nce = 21,443,656 Direct DSE = 24,992,574

Np = 21,403,543 U = 3.341

Ncp = 18,364,263 Census = 24,157,485

Error Source	Pos Gross Error	Neg Gross Error	Net Error	B(U)

Matching Error				
Mm	67,274	121,786	-54,511	0.34
	(23,216)	(25,173)	(32,640)	
Npm	72,090	60,142	11,948	
	(25,032)	(23,639)	(34,478)	
P Sample Collection				
Npa	91,449	82,514	8,934	-0.02
	(37,729)	(49,987)	(60,023)	
Ma	108,725	97,066	11,659	
	(74,071)	(51,135)	(91,590)	
P Sample Fabrication				
Npf		20,912	-20,912	0.22
		(20,912)	(20,912)	
Mf		60,413	-60,413	
		(60,413)	(60,413)	
E Sample Error				
Co	57,648	76,277	-18,629	-0.08
	(19,758)	(23,852)	(29,551)	
Cc	68,998	96,545	-27,547	-0.12
	(25,177)	(23,469)	(34,135)	
Model Bias (Tau)				
Imputation Error				
Ratio Estimator Bias			389	0.00
			(39)	
Net Sampling Error			0	
			(119,134)	

Moments of Error Components for
Evaluation Poststratum 15
Minority/non-owner/Large or
Medium MSA - Low

Nce = 6,310,050 Direct DSE = 7,803,395

Np = 7,660,305 U = 4.052

Ncp = 6,194,343 Census = 7,487,171

Error Source	Pos Gross Error	Neg Gross Error	Net Error	B(U)

Matching Error				
Mm	25,312	77,351	-52,039	0.81
	(6,795)	(30,179)	(27,662)	
Npm	20,110	19,728	382	
	(6,474)	(6,029)	(8,303)	
P Sample Collection				
Npa	29,114	22,751	6,363	0.33
	(15,940)	(9,852)	(14,494)	
Ma	13,095	28,921	-15,826	
	(9,066)	(10,768)	(8,021	
P Sample Fabrication				
Npf		1,172	-1,172	0.03
		(919)	(919)	
Mf		2,884	-2,884	
		(2,211)	(2,211)	
E Sample Error				
Co	60,172	41,682	18,491	0.28
	(27,583)	(15,084)	(24,118)	
Cc	36,168	42,828	-6,659	-0.10
	(15,700)	(12,730)	(15,593)	
Model Bias (Tau)				
Imputation Error				
Ratio Estimator Bias			1,506	0.02
			(151)	
Net Sampling Error			0	
			(65,577)	

Moments of Error Components for
Evaluation Poststratum 16
Minority/non-owner/All Other
TEAs

Nce = 8,229,779 Direct DSE = 9,718,222

Np = 8,386,177 U = 3.907

Ncp = 7,101,750 Census = 9,338,498

Error Source	Pos Gross Error	Neg Gross Error	Net Error	B(U)

Materials attached to these minutes were draft and preliminary material to inform the ESCAP Committee. The data and analysis contained in these documents are subject to revision and are not final. These materials report the results of research and analysis undertaken by Census Bureau staff. They have undergone a more limited review than official Census Bureau publications. Research results and conclusions expressed are those of the authors and do not necessarily indicate concurrence by the Census Bureau.

Loss Functions

Type of Loss Functions	Census Loss	A.C.E. Loss
1. Squared Error Loss	$\sum_i \left(Cen_i - T_i\right)^2$	$\sum_i \left(ACE_i - T_i\right)^2$
2. Weighted Squared Error Loss	$\sum_i \left(Cen_i - T_i\right)^2 / Cen_i$	$\sum_i \left(ACE_i - T_i\right)^2 / ACE_i$
3. Relative Squared Error Loss	$\sum_i \left(Cen_i - T_i\right)^2 / Cen_i^2$	$\sum_i \left(ACE_i - T_i\right)^2 / ACE_i^2$
4. Equal CD Squared Error Loss (Only for Districts)	$\sum_j Cen_j^2 \sum_i \left(Cen_i - T_i\right)^2$	$\sum_j Cen_j^2 \sum_i \left(ACE_i - T_i\right)^2$

Table A. U.S. Summary of Loss Functions

Geography	Distribution Method	DA Model	Correlation Bias Model	Total Census Population	Total ACE Actual Population	Total Sim ACE Population	Total Target Population	Weighted Levels	Weighted Share	Equal CD Share
State	GRODSE	N/A	No Correlation Bias	281,421,906	284,683,794	284,678,078	282,697,150	-48.1%	78.3%	N/A
		Model 1	Correlation Bias all Groups	281,421,906	284,683,794	284,678,078	283,408,878	109.9%	68.1%	N/A
		Model 1	Correlation Bias Black Only	281,421,906	284,683,794	284,678,078	283,444,371	128.7%	81.0%	N/A
		Model 1	Correlation Bias except NB 18-29, 0% Processing Error	281,421,906	284,683,794	284,678,078	285,088,512	1648.8%	12.5%	N/A
		Model 1	Correlation Bias except NB 18-29, 25% Processing Error	281,421,906	284,683,794	284,678,078	284,761,136	1756.5%	31.8%	N/A
		Model 1	Correlation Bias except NB 18-29, 50% Processing Error	281,421,906	284,683,794	284,678,078	284,434,836	1310.7%	50.0%	N/A
		Model 1	Correlation Bias except NB 18-29, 75% Processing Error	281,421,906	284,683,794	284,678,078	284,110,255	724.2%	65.6%	N/A
		Model 1	Correlation Bias except NB 18-29, 100% Processing Error	281,421,906	284,683,794	284,678,078	283,785,900	341.3%	78.0%	N/A
		Model 2	Correlation Bias except NB 18-29	281,421,906	284,683,794	284,678,078	283,782,669	339.3%	78.2%	N/A
		Model 2	Correlation Bias except NB 18-29, Revised DA	281,421,906	284,683,794	284,678,078	283,837,998	389.5%	79.3%	N/A
		Model 2	Correlation Bias except NB 18-29, All Owners, Revised DA	281,421,906	284,683,794	284,678,078	283,139,517	27.6%	90.8%	N/A
		Model 2	Correlation Bias except NB 18-29, Hisp same as Black, Revised DA	281,421,906	284,683,794	284,678,078	284,713,307	992.2%	132.6%	N/A

Table A. U.S. Summary of Loss Functions

Geography	Distribution Method	DA Model	Correlation Bias Model	Total Census Population	Total ACE Actual Population	Total Sim ACE Population	Total Target Population	Weighted Levels	Weighted Share	Equal CD Share
State	GRODSE	Model 2	Correlation Bias except NB 18-29, Hisp same as Black Corrected, Revised DA	281,421,906	284,683,794	284,678,078	284,191,619	562.7%	128.7%	N/A
		Model 2	Correlation Bias except NB 18-29, Hisp Renters same as Black, 90% Processing Error, Revised DA	281,421,906	284,683,794	284,678,078	284,457,602	1155.8%	129.6%	N/A
		Model 2	10% Correlation Bias Black Only, 90% Processing Error, Revised DA	281,421,906	284,683,794	284,678,078	282,900,999	-23.0%	76.1%	N/A
		Model 2	20% Correlation Bias Black Only, 90% Processing Error, Revised DA	281,421,906	284,683,794	284,678,078	282,975,113	-10.3%	79.2%	N/A
		Model 2	50% Correlation Bias Black Only, 90% Processing Error, Revised DA	281,421,906	284,683,794	284,678,078	283,198,131	41.6%	83.8%	N/A
		Model 2	75% Correlation Bias Black Only, 90% Processing Error, Revised DA	281,421,906	284,683,794	284,678,078	283,383,781	104.8%	82.1%	N/A
		Model 2	Fixed Odds Ratio Model except NB 18-29, Revised DA	281,421,906	284,683,794	284,678,078	283,840,365	412.1%	117.9%	N/A
		Model 2	Fixed Relative Risk Model except NB 18-29, Revised DA	281,421,906	284,683,794	284,678,078	283,838,451	402.3%	100.9%	N/A
		Model 2	Fixed Ratio of PM22 to PF22 Model except NB 18-29, Revised DA	281,421,906	284,683,794	284,678,078	283,841,734	403.3%	98.7%	N/A
		Model 2	Generalized Behavior Response Model except NB 18-29, Revised DA	281,421,906	284,683,794	284,678,078	283,835,552	389.7%	95.0%	N/A
		Model 2	Prithwis Das Gupta's Model except NB 18-29, Revised DA	281,421,906	284,683,794	284,678,078	283,838,808	378.5%	59.2%	N/A

Table A. U.S. Summary of Loss Functions

Geography	Distribution Method	DA Model	Correlation Bias Model	Total Census Population	Total ACE Actual Population	Total Sim ACE Population	Total Target Population	Weighted Levels	Weighted Share	Equal CD Share
State	GROSUC	N/A	No Correlation Bias	281,421,906	284,683,794	284,678,078	282,697,219	-54.7%	9.2%	N/A
		Model 1	Correlation Bias all Groups	281,421,906	284,683,794	284,678,078	283,412,224	92.5%	14.3%	N/A
		Model 1	Correlation Bias Black Only	281,421,906	284,683,794	284,678,078	283,447,040	109.2%	23.8%	N/A
		Model 1	Correlation Bias except NB 18-29, 0% Processing Error	281,421,906	284,683,794	284,678,078	285,091,168	1468.1%	-10.0%	N/A
		Model 1	Correlation Bias except NB 18-29, 25% Processing Error	281,421,906	284,683,794	284,678,078	284,763,567	1513.3%	-1.8%	N/A
		Model 1	Correlation Bias except NB 18-29, 50% Processing Error	281,421,906	284,683,794	284,678,078	284,437,366	1127.5%	6.5%	N/A
		Model 1	Correlation Bias except NB 18-29, 75% Processing Error	281,421,906	284,683,794	284,678,078	284,112,419	632.3%	13.9%	N/A
		Model 1	Correlation Bias except NB 18-29, 100% Processing Error	281,421,906	284,683,794	284,678,078	283,788,762	298.5%	20.1%	N/A
		Model 2	Correlation Bias except NB 18-29	281,421,906	284,683,794	284,678,078	283,785,652	296.8%	20.0%	N/A
		Model 2	Correlation Bias except NB 18-29, Revised DA	281,421,906	284,683,794	284,678,078	283,840,921	340.6%	20.8%	N/A
		Model 2	Correlation Bias except NB 18-29, All Owners, Revised DA	281,421,906	284,683,794	284,678,078	283,140,950	16.2%	21.3%	N/A
		Model 2	Correlation Bias except NB 18-29, Hisp same as Black, Revised DA	281,421,906	284,683,794	284,678,078	284,714,080	927.6%	99.2%	N/A

Table A. U.S. Summary of Loss Functions

Geography	Distribution Method	DA Model	Correlation Bias Model	Total Census Population	Total ACE Actual Population	Total Sim ACE Population	Total Target Population	Weighted Levels	Weighted Share	Equal CD Share
State	GROSUC	Model 2	Correlation Bias except NB 18-29, Hisp same as Black Corrected, Revised DA	281,421,906	284,683,794	284,678,078	284,192,056	525.1%	98.6%	N/A
		Model 2	Correlation Bias except NB 18-29, Hisp Renters same as Black, 90% Processing Error, Revised DA	281,421,906	284,683,794	284,678,078	284,457,965	1025.7%	78.2%	N/A
		Model 2	10% Correlation Bias Black Only, 90% Processing Error, Revised DA	281,421,906	284,683,794	284,678,078	282,901,001	-30.8%	8.2%	N/A
		Model 2	20% Correlation Bias Black Only, 90% Processing Error, Revised DA	281,421,906	284,683,794	284,678,078	282,975,624	-18.9%	10.5%	N/A
		Model 2	50% Correlation Bias Black Only, 90% Processing Error, Revised DA	281,421,906	284,683,794	284,678,078	283,199,460	29.7%	16.5%	N/A
		Model 2	75% Correlation Bias Black Only, 90% Processing Error, Revised DA	281,421,906	284,683,794	284,678,078	283,385,764	88.1%	19.9%	N/A
		Model 2	Fixed Odds Ratio Model except NB 18-29, Revised DA	281,421,906	284,683,794	284,678,078	283,842,386	371.1%	55.2%	N/A
		Model 2	Fixed Relative Risk Model except NB 18-29, Revised DA	281,421,906	284,683,794	284,678,078	283,840,762	360.6%	39.9%	N/A
		Model 2	Fixed Ratio of PM22 to PF22 Model except NB 18-29, Revised DA	281,421,906	284,683,794	284,678,078	283,844,642	365.8%	39.5%	N/A
		Model 2	Generalized Behavior Response Model except NB 18-29, Revised DA	281,421,906	284,683,794	284,678,078	283,836,410	355.5%	39.1%	N/A
		Model 2	Prithwis Das Gupta's Model except NB 18-29, Revised DA	281,421,906	284,683,794	284,678,078	283,842,090	328.1%	4.8%	N/A

Table A. U.S. Summary of Loss Functions

Geography	Distribution Method	DA Model	Correlation Bias Model	Total Census Population	Total ACE Actual Population	Total Sim ACE Population	Total Target Population	Weighted Levels	Weighted Share	Equal CD Share
Congressional District	GRODSE	N/A	No Correlation Bias	281,421,906	284,683,787	284,678,060	282,697,151	N/A	N/A	-0.5%
		Model 1	Correlation Bias all Groups	281,421,906	284,683,787	284,678,060	283,408,884	N/A	N/A	60.1%
		Model 1	Correlation Bias Black Only	281,421,906	284,683,787	284,678,060	283,444,376	N/A	N/A	69.2%
		Model 1	Correlation Bias except NB 18-29, 0% Processing Error	281,421,906	284,683,787	284,678,060	285,088,509	N/A	N/A	106.8%
		Model 1	Correlation Bias except NB 18-29, 25% Processing Error	281,421,906	284,683,787	284,678,060	284,761,146	N/A	N/A	97.5%
		Model 1	Correlation Bias except NB 18-29, 50% Processing Error	281,421,906	284,683,787	284,678,060	284,434,845	N/A	N/A	87.0%
		Model 1	Correlation Bias except NB 18-29, 75% Processing Error	281,421,906	284,683,787	284,678,060	284,110,248	N/A	N/A	75.9%
		Model 1	Correlation Bias except NB 18-29, 100% Processing Error	281,421,906	284,683,787	284,678,060	283,785,901	N/A	N/A	65.1%
		Model 2	Correlation Bias except NB 18-29	281,421,906	284,683,787	284,678,060	283,782,666	N/A	N/A	64.9%
		Model 2	Correlation Bias except NB 18-29, Revised DA	281,421,906	284,683,787	284,678,060	283,838,007	N/A	N/A	64.8%
		Model 2	Correlation Bias except NB 18-29, All Owners, Revised DA	281,421,906	284,683,787	284,678,060	283,139,516	N/A	N/A	57.8%
		Model 2	Correlation Bias except NB 18-29, Hisp same as Black, Revised DA	281,421,906	284,683,787	284,678,060	284,713,304	N/A	N/A	109.9%

Table A. U.S. Summary of Loss Functions

Geography	Distribution Method	DA Model	Correlation Bias Model	Total Census Population	Total ACE Actual Population	Total Sim ACE Population	Total Target Population	Weighted Levels	Weighted Share	Equal CD Share
Congressional District	GRODSE	Model 2	Correlation Bias except NB 18-29, Hisp same as Black Corrected, Revised DA	281,421,906	284,683,787	284,678,060	284,191,614	N/A	N/A	108.2%
		Model 2	Correlation Bias except NB 18-29, Hisp Renters same as Black, 90% Processing Error, Revised DA	281,421,906	284,683,787	284,678,060	284,457,601	N/A	N/A	109.4%
		Model 2	10% Correlation Bias Black Only, 90% Processing Error, Revised DA	281,421,906	284,683,787	284,678,060	282,901,223	N/A	N/A	14.7%
		Model 2	20% Correlation Bias Black Only, 90% Processing Error, Revised DA	281,421,906	284,683,787	284,678,060	282,975,111	N/A	N/A	26.5%
		Model 2	50% Correlation Bias Black Only, 90% Processing Error, Revised DA	281,421,906	284,683,787	284,678,060	283,198,142	N/A	N/A	55.4%
		Model 2	75% Correlation Bias Black Only, 90% Processing Error, Revised DA	281,421,906	284,683,787	284,678,060	283,383,776	N/A	N/A	68.6%
		Model 2	Fixed Odds Ratio Model except NB 18-29, Revised DA	281,421,906	284,683,787	284,678,060	283,840,362	N/A	N/A	81.5%
		Model 2	Fixed Relative Risk Model except NB 18-29, Revised DA	281,421,906	284,683,787	284,678,060	283,838,460	N/A	N/A	75.8%
		Model 2	Fixed Ratio of PM22 to PF22 Model except NB 18-29, Revised DA	281,421,906	284,683,787	284,678,060	283,841,749	N/A	N/A	74.6%
		Model 2	Generalized Behavior Response Model except NB 18-29, Revised DA	281,421,906	284,683,787	284,678,060	283,835,550	N/A	N/A	79.7%
		Model 2	Prithwis Das Gupta's Model except NB 18-29, Revised DA	281,421,906	284,683,787	284,678,060	283,838,809	N/A	N/A	56.1%

Table A. U.S. Summary of Loss Functions

Geography	Distribution Method	DA Model	Correlation Bias Model	Total Census Population	Total ACE Actual Population	Total Sim ACE Population	Total Target Population	Weighted Levels	Weighted Share	Equal CD Share
Congressional District	GROSUC	N/A	No Correlation Bias	281,421,906	284,683,787	284,678,060	282,697,215	N/A	N/A	-12.4%
		Model 1	Correlation Bias all Groups	281,421,906	284,683,787	284,678,060	283,412,231	N/A	N/A	50.8%
		Model 1	Correlation Bias Black Only	281,421,906	284,683,787	284,678,060	283,447,049	N/A	N/A	59.4%
		Model 1	Correlation Bias except NB 18-29, 0% Processing Error	281,421,906	284,683,787	284,678,060	285,091,163	N/A	N/A	97.1%
		Model 1	Correlation Bias except NB 18-29, 25% Processing Error	281,421,906	284,683,787	284,678,060	284,763,562	N/A	N/A	87.6%
		Model 1	Correlation Bias except NB 18-29, 50% Processing Error	281,421,906	284,683,787	284,678,060	284,437,367	N/A	N/A	76.9%
		Model 1	Correlation Bias except NB 18-29, 75% Processing Error	281,421,906	284,683,787	284,678,060	284,112,402	N/A	N/A	65.8%
		Model 1	Correlation Bias except NB 18-29	281,421,906	284,683,787	284,678,060	283,788,766	N/A	N/A	54.9%
		Model 2	Correlation Bias except NB 18-29	281,421,906	284,683,787	284,678,060	283,785,661	N/A	N/A	54.6%
		Model 2	Correlation Bias except NB 18-29, Revised DA	281,421,906	284,683,787	284,678,060	283,840,923	N/A	N/A	54.4%
		Model 2	Correlation Bias except NB 18-29, All Owners, Revised DA	281,421,906	284,683,787	284,678,060	283,140,960	N/A	N/A	44.4%
		Model 2	Correlation Bias except NB 18-29, Hisp same as Black, Revised DA	281,421,906	284,683,787	284,678,060	284,714,085	N/A	N/A	101.5%

Table A. U.S. Summary of Loss Functions

Geography	Distribution Method	DA Model	Correlation Bias Model	Total Census Population	Total ACE Actual Population	Total Sim ACE Population	Total Target Population	Weighted Levels	Weighted Share	Equal CD Share
Congressional District	GROSUC	Model 2	Correlation Bias except NB 18-29, Hisp same as Black Corrected, Revised DA	281,421,906	284,683,787	284,678,060	284,192,054	N/A	N/A	101.0%
		Model 2	Correlation Bias except NB 18-29, Hisp Renters same as Black, 90% Processing Error, Revised DA	281,421,906	284,683,787	284,678,060	284,457,962	N/A	N/A	97.8%
		Model 2	10% Correlation Bias Black Only, 90% Processing Error, Revised DA	281,421,906	284,683,787	284,678,060	282,901,231	N/A	N/A	2.1%
		Model 2	20% Correlation Bias Black Only, 90% Processing Error, Revised DA	281,421,906	284,683,787	284,678,060	282,975,620	N/A	N/A	13.3%
		Model 2	50% Correlation Bias Black Only, 90% Processing Error, Revised DA	281,421,906	284,683,787	284,678,060	283,199,462	N/A	N/A	42.2%
		Model 2	75% Correlation Bias Black Only, 90% Processing Error, Revised DA	281,421,906	284,683,787	284,678,060	283,385,764	N/A	N/A	56.9%
		Model 2	Fixed Odds Ratio Model except NB 18-29, Revised DA	281,421,906	284,683,787	284,678,060	283,842,377	N/A	N/A	71.4%
		Model 2	Fixed Relative Risk Model except NB 18-29, Revised DA	281,421,906	284,683,787	284,678,060	283,840,757	N/A	N/A	65.6%
		Model 2	Fixed Ratio of PM22 to PF22 Model except NB 18-29, Revised DA	281,421,906	284,683,787	284,678,060	283,844,643	N/A	N/A	67.4%
		Model 2	Generalized Behavior Response Model except NB 18-29, Revised DA	281,421,906	284,683,787	284,678,060	283,836,423	N/A	N/A	70.6%
		Model 2	Prithwis Das Gupta's Model except NB 18-29, Revised DA	281,421,906	284,683,787	284,678,060	283,842,095	N/A	N/A	46.3%

Table A. U.S. Summary of Loss Functions

Geography	Distribution Method	DA Model	Correlation Bias Model	Total Census Population	Total ACE Actual Population	Total Sim ACE Population	Total Target Population	Weighted Levels	Weighted Share	Equal Share
County	GRODSE	Model 2	Correlation Bias except NB 18-29, Revised DA	281,421,906	284,683,795	284,678,082	283,838,024	185.2%	N/A	87.3%
County	GROSUC	Model 2	Correlation Bias except NB 18-29, Revised DA	281,421,906	284,683,795	284,678,082	283,840,929	156.6%	N/A	81.6%

Materials attached to these minutes were draft and preliminary material to inform the ESCAP Committee. The data and analysis contained in these documents are subject to revision and are not final. These materials report the results of research and analysis undertaken by Census Bureau staff. They have undergone a more limited review than official Census Bureau publications. Research results and conclusions expressed are those of the authors and do not necessarily indicate concurrence by the Census Bureau.

Title 13-protected and/or other sensitive data, and/or detailed group quarters data that have not yet been officially released were deleted from the attached materials prior to their posting to this web page.

Table 1

- 2/22/2001

Sex Ratio by Single Year of Age: 2000 and 1990

Sex Ratios - 2000

Age	Total POP	Hispanic	Not Hispanic: Total	White	Black or African American	American Indian and Alaska Native	Asian and Pacific Islander
Total	96.3	105 9	95.0	95.8	90.4	96 6	94.5
0	105.0	104.8	105.0	105.6	102 9	104 8	104 9
1	104.6	104 1	104 7	105 4	103.1	103.9	100 8
2	104 7	104 6	104.8	105 3	103 5	103.2	101.5
3	104 5	104.6	104 5	105.3	102 9	103 1	100 4
4	105 0	104.7	105 0	105.8	103 1	104.0	101.6
5	105 0	104 8	105 1	105.7	103.1	103 8	103 7
6	104 9	104 5	105.0	105.7	102 8	101.8	105 1
7	105.1	104 4	105.2	105.8	103 4	101 4	104.7
8	104 7	104 3	104 8	105.3	103 0	102.6	105.8
9	105.1	104 6	105.1	105.7	103 0	102.7	106 5
10	105 2	104.5	105.3	105.9	103.1	103.9	106.0
11	105 0	104.7	105.1	105.6	102.8	102 0	106 7
12	105.0	104 3	105.1	105 6	102 8	104.3	105.5
13	105 0	104 0	105.1	105.7	103 0	104 3	105.1
14	105.4	105 4	105 4	106.0	103.0	104 6	105.9
15	105.7	106 6	105.5	106.0	103 8	103.6	105 5
16	106.3	110 4	105 6	106.1	104 2	103.1	104.5
17	107.0	115.1	105.5	105 9	104 3	104 5	104.7
18	105.4	117 2	103.2	103 8	100.6	103.7	103.5
19	104.3	119 4	101 6	102 3	97 8	103 2	102.1
20	104.7	122 2	101 4	102.3	97.3	102 7	102.1
21	104 8	123 3	101.3	102 6	95.8	103 6	100.9
22	104 6	123 4	100 9	102 4	94 4	101.8	100.0
23	104 3	121 3	100.8	102 7	93.0	102 6	-99.4
24	103.7	120 9	100 1	102.1	92.1	101.0	98 4

Sex Ratios - 1990

Age	Total POP	Hispanic	Not Hispanic: Total	White	Black or African American	American Indian and Alaska Native	Asian and Pacific Islander
Total	95.1	103.8	94.3	95.0	89.3	97.1	95.7
0	104.7	103.9	104.8	105.5	102.0	102 8	104 8
1	104.8	104.4	104.8	105.4	102.2	103.4	105 8
2	104.8	104.5	104.9	105.5	102.3	104 5	104 1
3	104.8	104.3	104.8	105.5	102.1	104.4	104 3
4	105.0	104.3	105.0	105.7	102.3	103.5	103.3
5	104.8	103.9	105.0	105.6	102.5	102.9	102.7
6	104 9	104.1	105.0	105.6	102.4	103 4	102 3
7	104.7	103 6	104.8	105.5	102.1	102 6	102.9
8	104.8	104 4	104.9	105.6	102.2	102.5	102 9
9	105.2	104.6	105.2	106.0	102.4	103.1	103 4
10	105.3	104.8	105.3	106.1	102.4	103 9	103.5
11	105 0	105 2	104 8	105.8	101.5	104.4	103.6
12	104.8	103.9	104.9	105 7	101.8	102 4	103.2
13	104 8	104.0	104.9	105.6	101.6	105.4	104.3
14	105.2	104 9	105 3	105.9	102.8	103 5	105 5
15	105 6	106.1	105.5	106.0	102.8	105 5	107 4
16	106 0	109.2	105 6	105.9	103.6	103.8	107 3
17	106.3	114.2	105.3	105.6	103.1	106 8	106.9
18	104 9	116.4	103 5	103 9	100 8	106 9	106 9
19	104 2	118.5	102.5	103.1	98.4	106 2	107.5
20	104.4	120 9	102.4	103 0	97 5	106 5	108 9
21	104 5	122.7	102.2	103.0	96.6	106 3	107 8
22	104 1	121.6	101.9	103.1	94 6	105 0	106 5
23	103.1	120.3	100.9	102.3	93.1	101 3	—105 2
24	102.7	119.6	100.5	102.1	92.0	101 8	103.2

Table 1 (cont)

2/22/2001

* Sex Ratio by Single Year of Age: 2000 and 1990

Age	Sex Ratios - 2000		Not Hispanic					Sex Ratios - 1990		Not Hispanic				
	Total POP	Hispanic	Total	White	Black or African American	American Indian and Alaska Native	Asian and Pacific Islander	Total POP	Hispanic	Total	White	Black or African American	American Indian and Alaska Native	Asian and Pacific Islander
25	103.7	120.4	100.2	102.3	91.6	101.6	98.4	101.7	117.1	99.8	101.4	90.9	101.1	100.8
26	102.5	117.5	99.4	101.6	90.6	99.9	97.4	100.9	114.8	99.3	100.9	90.5	101.0	98.5
27	102.1	117.2	99.1	101.2	90.8	101.1	97.0	100.8	114.0	99.2	100.9	90.0	99.8	98.5
28	101.9	115.4	99.3	101.5	90.3	100.1	96.6	100.2	112.9	98.8	100.6	89.2	96.9	96.8
29	101.3	115.1	98.9	101.1	89.2	99.0	96.5	100.2	112.8	98.8	100.9	88.4	95.2	95.8
30	102.8	117.0	100.3	102.5	91.0	99.3	96.9	99.5	110.2	98.4	100.4	87.8	95.7	94.2
31	101.4	114.6	99.2	101.3	90.2	101.1	95.1	99.0	108.8	98.0	100.2	87.0	93.9	93.3
32	101.2	114.2	99.0	101.4	89.2	98.5	94.4	99.0	107.9	98.1	100.2	87.1	94.9	93.6
33	100.5	113.0	98.4	100.6	89.0	96.4	95.6	98.3	107.0	97.5	99.8	85.9	91.9	91.4
34	100.6	113.5	98.6	100.8	89.5	97.9	95.4	99.2	107.0	98.5	100.9	86.7	92.5	91.6
35	100.7	112.5	98.9	101.1	89.3	94.8	96.3	98.8	105.4	98.2	100.4	86.6	93.3	91.3
36	99.3	109.3	97.9	100.0	88.3	96.0	94.6	98.0	103.5	97.5	99.8	85.9	91.7	90.1
37	99.3	108.1	98.1	100.0	88.8	95.7	95.7	97.9	103.1	97.5	99.9	85.1	92.3	88.5
38	99.2	109.0	98.0	100.1	87.9	95.2	93.9	97.7	102.3	97.3	99.7	84.8	92.7	87.9
39	98.5	106.5	97.6	99.7	88.5	91.5	91.3	98.5	103.0	98.2	100.6	86.5	94.6	88.2
40	99.7	108.2	98.7	100.9	89.3	92.9	93.1	97.6	100.4	97.4	99.8	85.6	94.9	86.7
41	98.2	104.7	97.5	99.6	88.0	91.9	91.0	97.1	99.5	96.9	99.2	84.8	93.1	86.7
42	98.5	104.8	97.8	99.8	89.0	92.4	91.7	97.3	98.7	97.2	99.3	85.1	94.5	86.5
43	97.7	102.7	97.2	99.4	87.7	91.4	88.8	97.1	97.7	97.1	99.1	84.6	94.4	86.4
44	97.6	102.3	97.1	99.4	87.6	91.0	87.5	97.6	99.3	97.5	99.5	85.9	92.7	90.7
45	98.2	103.0	97.7	100.0	88.1	92.5	88.6	96.9	97.3	96.9	98.9	84.4	93.5	90.2
46	96.7	100.0	96.4	98.5	86.7	90.8	87.4	96.4	95.2	96.5	98.3	83.7	93.1	91.9
47	96.7	98.6	96.5	98.8	86.2	91.8	87.0	96.2	94.6	96.3	98.1	83.3	94.5	92.6
48	96.7	98.9	96.5	98.8	85.2	90.5	86.6	95.3	94.8	95.3	97.1	82.5	94.1	93.2
49	96.2	97.9	96.1	98.3	86.3	93.8	84.9	96.2	96.3	96.2	97.9	83.9	98.3	97.2
50	96.6	97.4	96.6	98.9	86.5	92.3	85.9	95.3	94.3	95.4	97.2	82.7	91.4	97.1
51	95.8	95.5	95.8	98.0	85.2	93.6	85.9	94.9	93.0	95.1	96.7	82.2	92.6	96.9
52	96.0	95.1	96.1	98.0	85.2	92.0	87.0	94.4	92.6	94.6	96.3	81.7	91.9	95.1

Table 1 (cont)

2/22/2001

Sex Ratio by Single Year of Age: 2000 and 1990

	Sex Ratios - 2000							Sex Ratios - 1990						
			Not Hispanic							Not Hispanic				
Age	Total POP	Hispanic	Total	White	Black or African American	American Indian and Alaska Native	Asian and Pacific Islander	Total POP	Hispanic	Total	White	Black or African American	American Indian and Alaska Native	Asian and Pacific Islander
53	95.8	94.4	95.9	97.7	84.9	93.7	86.4	93.8	91.7	93.9	95.7	80.6	94.2	93.9
54	95.0	93.7	95.1	97.1	83.5	92.3	88.7	93.3	91.4	93.5	95.3	80.9	93.0	90.3
55	94.7	93.5	94.9	96.8	83.2	90.6	88.4	92.7	90.9	92.9	94.7	80.5	92.0	88.3
56	94.0	90.4	94.3	96.1	82.8	92.6	89.1	92.1	89.7	92.3	94.3	79.2	91.0	84.3
57	93.5	88.6	93.9	95.6	81.8	92.3	89.4	91.2	88.1	91.4	93.4	79.3	88.3	80.7
58	92.8	90.3	93.0	94.8	80.4	94.3	89.0	90.5	88.9	90.6	92.5	77.8	89.1	78.5
59	92.1	89.4	92.3	94.1	80.5	93.4	89.5	90.5	88.4	90.6	92.4	78.3	91.7	78.0
60	92.0	88.6	92.3	94.0	80.1	91.9	90.8	89.2	87.4	89.3	91.1	76.9	89.8	77.1
61	91.6	87.4	91.9	93.5	79.7	93.8	91.3	88.3	86.8	88.4	90.2	75.9	89.3	74.8
62	90.9	86.6	91.2	93.0	79.2	92.0	88.2	87.5	85.1	87.6	89.3	75.5	88.9	75.1
63	89.8	84.7	90.1	91.8	78.3	92.2	87.8	86.3	83.6	86.4	88.0	74.9	88.2	74.9
64	88.6	83.1	89.0	90.9	76.8	89.3	82.6	84.3	81.6	84.4	85.8	74.2	84.7	75.0
65	88.1	83.4	88.4	90.2	76.6	88.1	82.9	83.0	82.0	83.0	84.2	73.3	86.3	77.6
66	86.7	81.6	87.1	89.1	74.3	87.3	79.3	82.2	84.9	82.1	83.1	73.3	80.5	80.7
67	85.5	79.7	85.9	87.8	73.5	84.3	77.5	81.2	81.4	81.2	82.1	72.6	84.6	81.7
68	84.9	80.6	85.2	87.0	72.3	82.8	76.7	80.1	80.3	80.1	80.9	71.5	80.7	83.6
69	83.3	79.2	83.5	85.3	71.4	85.2	73.8	78.7	75.9	78.8	79.6	70.7	77.0	81.4
70	81.9	78.0	82.2	84.0	68.9	84.7	74.8	77.1	76.0	77.2	78.0	68.9	77.2	81.8
71	80.7	76.4	80.9	82.6	67.8	81.6	73.3	75.7	72.7	75.9	76.6	67.6	75.3	80.6
72	79.1	76.1	79.2	80.8	66.1	79.2	73.7	74.4	71.4	74.6	75.3	65.3	76.2	79.3
73	77.2	74.6	77.3	78.7	65.3	77.0	72.6	72.8	69.4	72.7	73.4	63.4	72.7	80.8
74	74.7	72.8	74.8	76.1	63.3	73.8	71.5	70.5	68.0	70.6	71.2	62.4	70.3	79.9
75	72.8	71.2	72.9	73.8	62.9	73.4	75.4	68.5	67.4	68.6	69.1	61.2	65.2	80.8
76	71.3	74.0	71.2	71.9	61.8	72.1	76.2	66.6	64.4	66.7	67.2	59.8	65.6	81.8
77	69.7	72.9	69.5	70.2	60.1	69.5	76.9	64.1	62.9	64.1	64.4	58.0	70.2	84.9
78	68.0	69.7	67.9	68.5	58.9	67.0	77.2	62.2	62.7	62.2	62.3	58.1	65.2	88.4
79	65.8	65.2	65.8	66.5	56.6	62.9	72.6	58.9	60.8	58.8	58.7	55.6	60.8	94.0
80	63.1	64.7	63.0	63.6	54.0	59.5	73.4	57.2	61.8	57.0	56.8	55.0	61.4	95.0

Table 1 (cont)

2/22/2001

Sex Ratio by Single Year of Age: 2000 and 1990

Age	Sex Ratios - 2000 Total POP	Hispanic	Not Hispanic Total	White	Black or African American	American Indian and Alaska Native	Asian and Pacific Islander	Sex Ratios - 1990 Total POP	Hispanic	Not Hispanic Total	White	Black or African American	American Indian and Alaska Native	Asian and Pacific Islander
81	61.1	62.2	61.0	61.5	52.9	62.5	72.7	55.1	60.9	54.9	54.6	53.4	58.9	97.0
82	59.2	60.8	59.1	59.7	49.5	55.8	67.9	52.9	61.0	52.6	52.3	51.5	60.9	94.9
83	56.1	57.4	56.0	56.5	46.6	54.0	69.3	50.4	60.7	50.1	49.7	49.1	59.0	95.2
84	53.5	56.4	53.4	53.8	45.8	51.5	65.3	47.9	58.1	47.6	47.1	48.2	57.9	90.6
85	50.5	53.5	50.3	50.7	43.2	51.2	66.7	45.9	57.6	45.5	44.9	48.3	59.5	87.7
86	47.7	51.2	47.5	47.8	40.8	44.8	66.0	43.8	59.1	43.4	42.8	45.9	57.0	81.9
87	45.2	49.8	45.0	45.2	39.1	49.4	65.9	41.7	56.1	41.3	40.6	45.1	54.2	79.4
88	42.7	48.6	42.5	42.4	38.8	48.0	68.6	40.2	59.2	39.7	39.1	43.3	62.8	73.5
89	39.5	47.5	39.2	39.0	37.0	42.6	68.7	37.5	54.0	37.1	36.2	42.3	54.0	67.0
90	37.6	47.7	37.3	36.9	36.0	48.0	69.2	35.5	54.2	35.0	34.4	39.7	61.7	66.0
91	34.9	46.2	34.6	34.1	33.9	44.2	70.9	33.0	50.1	32.6	32.1	38.0	48.9	54.9
92	33.2	47.0	32.8	32.2	34.1	42.1	66.7	31.9	48.6	31.5	31.0	35.6	45.7	50.4
93	30.7	45.8	30.2	29.7	31.1	32.1	62.0	30.5	46.1	30.1	29.5	35.5	57.9	49.7
94	29.1	44.1	28.6	27.9	30.3	43.9	65.4	29.2	44.8	28.8	28.3	34.3	38.5	38.2
95	27.4	43.4	26.8	26.1	30.0	41.0	58.9	26.5	42.6	28.1	27.6	33.7	61.3	32.6
96	25.4	43.5	24.8	24.1	27.5	42.9	48.9	27.2	45.7	26.8	26.3	32.7	33.7	32.2
97	24.2	46.1	23.5	22.6	26.7	45.6	51.3	26.4	41.2	26.0	25.2	32.1	58.1	37.6
98	23.3	43.2	22.6	21.4	28.7	29.2	48.1	25.3	39.7	24.9	23.9	33.2	48.3	45.4
99	23.7	46.0	22.9	21.2	30.1	47.1	52.7	27.9	55.0	27.1	25.7	35.6	55.0	53.2
100	36.4	60.7	34.7	32.6	38.5	71.6	50.6	27.4	55.3	26.3	23.6	37.1	60.0	52.8

filename - sexratio20001990

Table 2

2/22/2001

Household Population Sex Ratio by Single Year of Age: 2000 and 1990

	Sex Ratios - 2000		Not Hispanic					Sex Ratios - 1990		Not Hispanic				
Age	Total POP	Hispanic	Total	White	Black or African American	American Indian and Alaska Native	Asian and Pacific Islander	Total POP	Hispanic	Total	White	Black or African American	American Indian and Alaska Native	Asian and Pacific Islander
Total	95.3	103.6	94.2	95.7	85.6	94.3	93.9	94.3	101.3	93.6	94.9	85.5	94.7	95.0
0	105.0	104.8	105.0	105.5	102.9	104.6	104.9	104.7	104.0	104.8	105.5	102.0	102.8	104.8
1	104.6	104.1	104.7	105.4	103.0	103.9	100.8	104.8	104.4	104.8	105.4	102.2	103.4	105.7
2	104.7	104.6	104.8	105.3	103.5	103.2	101.5	104.8	104.5	104.9	105.5	102.3	104.5	104.1
3	104.5	104.6	104.5	105.3	102.9	103.1	100.4	104.6	104.3	104.6	105.5	102.1	104.5	104.2
4	105.0	104.7	105.0	105.6	103.1	104.0	101.6	104.9	104.3	105.0	105.7	102.3	103.5	103.3
5	105.0	104.8	105.1	105.7	103.1	103.6	103.7	104.6	103.9	104.9	105.6	102.5	103.0	102.7
6	104.9	104.5	105.0	105.7	102.8	101.8	105.1	104.8	104.1	104.9	105.6	102.3	103.4	102.3
7	105.0	104.4	105.2	105.6	103.4	101.4	104.6	104.7	103.6	104.8	105.5	102.1	102.6	102.9
8	104.7	104.3	104.8	105.3	102.9	102.5	105.8	104.8	104.4	104.8	105.5	102.1	102.6	102.8
9	105.0	104.6	105.1	105.6	102.9	102.6	106.4	105.1	104.5	105.2	105.9	102.3	103.0	103.3
10	105.1	104.5	105.2	105.6	103.0	103.9	106.0	105.2	104.6	105.2	106.0	102.2	103.8	103.4
11	104.9	104.6	105.0	105.5	102.6	101.8	106.7	104.8	105.1	104.8	105.6	101.3	104.2	103.7
12	104.8	104.2	104.9	105.5	102.4	104.2	105.5	104.6	103.8	104.7	105.5	101.5	102.1	103.2
13	104.7	103.7	104.9	105.5	102.4	104.0	105.1	104.5	103.7	104.6	105.4	101.0	104.9	104.3
14	105.0	104.9	105.0	105.7	101.6	104.0	105.7	104.9	104.3	104.9	105.7	101.6	103.1	105.4
15	104.8	105.6	104.7	105.5	101.5	102.6	105.1	104.8	105.0	104.8	105.6	100.7	104.5	107.1
16	105.0	108.6	104.4	105.3	100.5	101.5	104.0	104.8	107.3	104.5	105.3	100.3	102.0	106.9
17	105.2	112.7	103.9	104.9	99.6	102.4	103.9	104.9	111.6	104.1	105.0	99.3	104.3	106.2
18	107.7	115.7	106.0	108.0	98.0	102.6	106.4	.105.0	113.1	103.9	105.2	97.4	104.0	107.6
19	103.7	116.4	101.0	103.1	92.0	99.6	103.5	101.4	113.4	99.8	101.3	91.2	99.9	106.5
20	101.0	117.5	97.5	99.6	87.2	96.3	100.9	99.3	114.4	97.3	99.0	87.0	98.0	106.5
21	99.9	117.6	96.2	98.9	83.6	96.6	99.6	98.7	115.6	96.4	98.2	85.1	97.3	105.1
22	98.9	117.6	95.0	98.0	81.1	94.9	97.8	98.2	114.9	96.0	98.2	83.0	97.3	103.3
23	99.1	115.9	95.5	99.0	79.9	96.2	97.1	98.0	114.0	96.0	98.4	82.1	94.7	102.2
24	99.1	116.0	95.5	99.1	79.6	94.9	96.3	98.3	113.5	96.4	99.0	81.6	95.4	100.7

Table 2 (cont)

2/22/2001

Household Population Sex Ratio by Single Year of Age: 2000 and 1990

	Sex Ratios - 2000							Sex Ratios - 1990						
			Not Hispanic							Not Hispanic				
Age	Total POP	Hispanic	Total	White	Black or African American	American Indian and Alaska Native	Asian and Pacific Islander	Total POP	Hispanic	Total	White	Black or African American	American Indian and Alaska Native	Asian and Pacific Islander
25	99.6	115.8	96.2	99.8	79.9	96.5	96.7	97.8	111.4	96.2	98.9	81.0	95.4	98.6
26	98.6	113.0	95.6	99.4	78.8	94.6	95.9	97.4	109.3	96.0	98.7	81.2	96.1	96.6
27	98.4	112.8	95.5	99.1	79.7	96.3	95.7	97.5	108.7	96.2	98.9	81.0	94.7	96.7
28	98.4	111.3	96.0	99.6	79.4	95.2	95.4	97.0	107.5	95.9	98.7	79.9	92.3	95.2
29	97.9	110.9	95.7	99.2	78.7	94.4	95.4	97.3	107.6	96.1	99.2	79.9	90.8	94.6
30	99.6	113.0	97.3	100.7	81.1	94.9	96.0	96.7	105.2	95.9	98.9	79.4	91.5	93.2
31	98.2	110.5	96.1	99.5	79.9	96.6	94.1	96.4	104.1	95.7	98.7	78.9	89.6	92.4
32	98.1	110.2	96.0	99.6	79.2	94.2	93.6	96.5	103.4	95.8	98.8	79.3	90.8	92.7
33	97.3	109.0	95.4	98.9	79.0	91.9	94.7	96.0	102.6	95.4	98.4	78.5	88.1	90.7
34	97.6	109.6	95.7	99.1	79.8	93.5	94.5	97.0	102.7	96.4	99.6	79.7	89.0	90.9
35	97.8	108.7	96.2	99.4	80.2	90.8	95.5	96.6	101.2	96.2	99.2	79.7	89.7	90.6
36	96.4	105.7	95.1	98.3	79.2	91.4	93.8	96.0	99.6	95.7	98.6	79.5	88.5	89.5
37	96.4	104.4	95.3	98.3	79.6	91.8	94.9	96.1	99.4	95.8	98.8	79.1	89.4	87.9
38	96.5	105.4	95.3	98.5	79.0	91.3	93.1	95.8	98.7	95.6	98.6	78.9	89.7	87.3
39	95.9	103.0	95.0	98.1	80.0	87.8	90.7	96.8	99.5	96.6	99.5	81.2	91.3	87.7
40	97.3	104.8	96.3	99.4	81.3	89.1	92.4	96.0	97.1	95.9	98.8	80.6	91.6	86.3
41	95.8	101.4	95.2	98.2	80.0	88.3	90.3	95.7	96.5	95.6	98.3	80.4	90.4	86.4
42	96.3	101.7	95.7	98.5	81.4	88.9	91.1	96.0	96.0	96.0	98.5	81.0	92.4	86.2
43	95.6	99.6	95.1	98.1	80.4	88.3	88.3	96.0	95.0	96.0	98.3	80.9	91.9	86.1
44	95.6	99.3	95.2	98.3	80.7	88.3	87.0	96.5	96.6	96.5	98.8	82.4	90.1	90.4
45	96.4	100.3	96.0	98.9	81.9	89.7	88.1	95.8	94.9	95.9	98.2	81.3	91.0	89.9
46	95.0	97.2	94.7	97.5	80.7	88.1	87.0	95.5	93.1	95.7	97.7	80.9	91.1	91.7
47	95.1	96.1	95.0	97.8	80.8	89.4	86.6	95.3	92.4	95.5	97.5	80.7	92.4	92.4
48	95.2	96.5	95.0	97.9	80.3	88.2	86.3	94.4	92.7	94.5	96.5	80.0	91.9	93.0
49	94.8	95.7	94.7	97.5	81.6	91.3	84.4	95.4	94.1	95.5	97.4	81.5	96.0	96.6
50	95.4	95.3	95.4	98.1	82.4	90.2	85.6	94.5	92.2	94.7	96.6	80.4	89.8	96.8
51	94.6	93.6	94.7	97.2	81.2	91.5	85.6	94.2	91.3	94.4	96.3	80.4	90.7	96.8
52	95.0	93.5	95.1	97.4	81.7	89.8	86.6	93.8	91.0	94.0	95.9	79.9	90.5	95.0

Table 2 (cont)

2/22/2001

Household Population Sex Ratio by Single Year of Age: 2000 and 1990

	Sex Ratios - 2000							Sex Ratios - 1990						
			Not Hispanic							Not Hispanic				
Age	Total POP	Hispanic	Total	White	Black or African American	American Indian and Alaska Native	Asian and Pacific Islander	Total POP	Hispanic	Total	White	Black or African American	American Indian and Alaska Native	Asian and Pacific Islander
53	94.9	92 6	95 0	97.1	81.5	92 0	86.2	93.2	90.2	93.4	95.3	78.9	92.4	93.7
54	94.0	92.1	94.2	96.5	80.4	90.4	88.5	92.7	89.6	93.0	94.9	79.3	91.5	90.1
55	93.9	92.0	94 0	96.3	80.5	89 3	88.2	92.1	89.4	92.3	94.3	78.9	90.3	88.1
56	93.3	89 0	93.7	95.6	80 4	91.0	89.0	91.6	88.5	91.8	93.9	77.8	89 8	84.1
57	92.9	87.3	93.3	95 2	79.5	90 8	89.2	90.7	86.9	91.0	93.1	77 8	87.1	80.6
58	92.2	89 0	92 4	94 4	78.3	92 8	88.9	90.1	87.7	90.2	92.2	76.5	87.6	78.3
59	91 5	88 2	91.7	93 6	78.4	91 9	89.4	90.0	87 2	90.1	92 1	76.9	90 0	77.7
60	91 4	87 5	91.7	93.6	78.2	90 8	90.7	88.8	86.2	88 9	90.8	75 6	88 5	76 9
61	91.0	86 2	91 4	93.1	77.9	92 8	91.1	87.9	85 9	88.0	89.9	74.7	88.2	74 7
62	90.4	85 6	90.8	92.7	77.7	91 0	88.0	87.1	84.3	87.3	89.1	74.3	88.0	74.8
63	89 3	83 9	89.7	91 5	76 8	90.7	87 7	86.0	82.8	86.1	87.8	73 8	86.8	74 7
64	88.2	82 2	88.6	90.7	75 3	88 4	82.5	84.0	81.0	84.2	85.6	73 2	83.5	74.8
65	87.7	82 6	88 1	90.0	75 3	87.0	82.7	82.7	81.2	82.8	84.0	72.3	85.5	77.4
66	86.4	80.7	86 8	89 0	72.9	86.4	79.2	82 0	84.3	81.9	83 0	72 4	79.9	80 6
67	85 2	78 9	85.7	87 8	72 2	83.8	77 4	81.0	80.8	81.1	82.0	71 7	83.2	81 5
68	84 7	80.1	85.0	87.0	71.1	82 2	76.7	80.0	79 8	80.0	80.9	70 7	79.9	83 3
69	83.2	78 7	83 4	85.3	70.2	84 2	73.6	78.7	75.3	78.8	79.7	69.8	75 8	81.0
70	81.8	77 4	82.1	84.0	67 8	84.3	74 6	77.1	75 5	77.2	78.0	68 1	76 5	81.7
71	80.7	75 8	81 0	82 8	65 7	80 6	73.1	75 8	72.3	76.0	76.8	68 9	74.5	80.5
72	79 2	75 6	79.4	81.1	65 1	78 9	73 6	74.6	71.0	74.8	75.6	64 8	75.2	79.2
73	77.4	74 2	77.6	79.1	64 4	76 5	72.6	72.9	69.2	73 0	73.8	63.1	72 3	80.9
74	75 0	72.4	75.2	76 6	62.7	73.7	71 4	71.0	67.6	71.1	71.8	62.1	70.4	79 7
75	73 2	70 8	73 4	74 4	62 0	72 6	75 4	69 1	67.3	69 2	69.8	60.8	64.6	80 9
76	71 9	73 6	71.8	72 6	61 2	71 7	76 2	67.4	64.2	67.5	68.1	59.6	65.1	81.9
77	70 5	72 9	70 4	71 2	59 7	69.0	76 8	65.1	62.8	65.1	65.6	57.8	69 2	84.9
78	69 0	69 6	69 0	69 7	58 7	66 5	77 6	63 4	62 7	63.5	63.6	58.0	64.8	89 0
79	67.2	65 2	67.3	68.1	56.4	63 0	73 1	60.4	60.9	60 4	60.4	55.7	60.1	94 6
80	64.7	64 6	64.7	65.5	53.8	60 0	73.6	59.0	61.9	58.9	58 8	55 2	61 6	96 2

Table 2 (cont)

2/22/2001

Household Population Sex Ratio by Single Year of Age: 2000 and 1990

Age	Sex Ratios - 2000: Total POP	Hispanic	Not Hispanic: Total	White	Black or African American	American Indian and Alaska Native	Asian and Pacific Islander	Sex Ratios - 1990: Total POP	Hispanic	Not Hispanic: Total	White	Black or African American	American Indian and Alaska Native	Asian and Pacific Islander
81	63.1	62.4	63.2	63.8	53.2	62.2	73.4	57.4	61.3	57.3	57.1	53.9	58.9	98.6
82	61.6	60.9	61.6	62.4	50.2	55.7	68.9	55.5	61.4	55.3	55.2	52.1	60.4	96.7
83	58.9	58.2	58.9	59.7	47.4	54.2	69.9	53.4	61.2	53.2	53.0	50.0	59.5	96.9
84	56.7	57.2	56.7	57.3	46.8	52.0	66.6	51.4	59.0	51.1	50.8	49.3	58.3	93.4
85	53.9	54.7	53.9	54.5	44.3	51.5	68.8	49.9	58.9	49.5	49.1	49.7	58.4	91.4
86	51.5	52.0	51.6	52.0	42.2	44.2	67.6	48.4	60.5	48.0	47.6	47.6	57.4	85.1
87	49.4	51.0	49.4	49.9	40.5	49.1	68.2	46.6	57.5	46.3	45.8	46.8	54.1	84.0
88	47.6	50.6	47.5	47.6	40.7	49.7	71.8	45.7	60.7	45.2	44.8	45.5	53.8	79.3
89	44.3	49.3	44.1	44.1	39.0	43.1	71.8	43.3	55.7	42.8	42.3	44.6	54.2	71.9
90	43.2	49.1	43.0	42.9	38.2	50.2	74.1	41.4	56.7	40.9	40.4	42.4	62.2	70.2
91	41.0	49.4	40.7	40.5	36.4	46.5	76.3	39.4	51.3	39.0	38.7	40.4	51.2	58.6
92	39.5	49.7	39.1	38.8	37.3	43.2	71.1	38.9	51.8	38.6	38.3	39.7	46.0	57.3
93	37.2	48.3	36.7	36.4	34.6	33.0	68.2	37.7	47.7	37.4	36.9	39.0	66.2	57.7
94	35.9	46.4	35.4	35.0	33.5	43.3	73.1	36.7	47.0	36.4	36.1	37.7	41.3	43.2
95	34.6	47.3	34.0	33.4	33.2	47.1	65.1	36.4	43.6	36.2	35.8	37.8	74.1	37.4
96	32.3	46.2	31.7	31.3	30.8	40.5	53.0	35.6	48.1	35.1	34.9	37.1	30.3	37.5
97	31.8	50.7	30.8	30.2	29.7	48.3	58.1	35.1	43.1	34.8	34.4	35.3	64.3	45.3
98	31.1	46.2	30.3	29.3	32.6	31.4	52.6	35.2	44.2	34.8	33.7	39.5	64.3	48.5
99	31.9	49.8	30.8	29.3	33.3	51.8	64.3	38.2	61.6	37.0	35.9	40.3	59.2	63.6
100	36.4	60.7	34.7	32.6	38.5	71.6	50.6	40.4	61.7	38.9	36.4	43.7	68.7	66.0

filename - sexratiohshd20001990

Table 3

2/22/2001

Group Quarters Population Sex Ratio by Single Year of Age and Race (IMPRACE)

			Sex Ratios - 2000				
			Not Hispanic				
Age	Total POP	Hispanic	Total	White	Black or African American	American Indian and Alaska Native	Asian and Pacific Islander
Total	137.4	320.5	127.4	100.1	273 3	216.0	122.3
0	109.8	105 6	111 3	114 4	107.2	104 5	134.6
1	107.3	102.6	109 4	109.6	108.2	127 2	108.1
2	106.2	107 6	105.6	109 0	103.9	114 1	91.7
3	108 1	114.5	105.3	107.4	103.9	97 2	108 0
4	106 1	109.6	104.5	102.3	106 3	100.0	110.5
5	114 5	102 4	120.5	127.8	115.2	103.0	121.2
6	115 9	110.0	118.5	124.3	117.0	108.7	88.3
7	120.2	108 3	125 2	142 2	110.6	105 2	128.4
8	126 2	104 9	134 4	143.6	129.2	133.0	93 5
9	136.1	116 9	142 6	157.0	129 6	124.0	134 1
10	158 5	130 7	167.2	186.5	156.4	116.1	112.3
11	167.2	145 6	172.6	194.1	153.8	141 0	123.3
12	175.2	149 2	181 0	196 0	176.2	117.4	116 8
13	186.4	163 8	191 6	196.9	197 7	131 9	129.5
14	200 8	199.0	201.2	192.0	225 5	149 0	172.6
15	235.5	254 4	232 0	205.7	289 1	157.6	226.6
16	273 2	315 9	265 3	237.9	327 3	170 6	200.8
17	285 0	386 3	268 6	240.1	339.8	184.2	175.2
18	89 5	149 2	86 0	79.0	119 4	114.7	89.8
19	106.6	164 9	103 6	89.9	127.3	131.1	97.8
20	130 0	231.3	124 6	116.7	170.2	176 6	108.8
21	158 5	301 5	150 0	137.4	219 9	216.4	111 2
22	231 1	422 3	216.5	194.3	327 7	267 9	133.7
23	384.9	601 5	359 9	330.9	488 9	333 2	171 3
24	470 4	681.6	440.9	392.1	624.3	374.8	197.1

			Sex Ratios - 1990				
			Not Hispanic				
Age	Total POP	Hispanic	Total	White	Black or African American	American Indian and Alaska Native	Asian and Pacific Islander
Total	130 2	331.7	121.6	101.2	256.3	239 5	135 0
0	104.9	98.1	107 4	114.3	101.8	107 8	87.5
1	107.3	105.9	107.8	108.0	107.9	86.7	118.7
2	105.4	102.0	106.9	109 0	106.1	90 0	98.8
3	109 3	105.9	110.6	113 4	106.4	96.3	128.1
4	110 6	105 7	112.5	110.6	119.3	90.4	94 9
5	114.3	106.1	117 0	121.9	112.4	97.7	99 0
6	124.1	113.7	127.5	127.2	131.2	113.5	109.6
7	128.3	101.9	136.9	146.3	127.8	104 8	95 5
8	142.9	118.0	149 7	154.5	147.0	100.0	133.3
9	152.4	134.1	156.9	162.0	150.7	148 4	125.0
10	167.2	131.1	175.2	177.8	177.7	123.3	142.0
11	177 9	145.3	184.5	194.5	178.1	143.2	91.0
12	174.7	152 2	178 8	180.9	183.3	147 6	113 6
13	176.3	183 4	175.2	164.8	206.9	144.4	136.5
14	176.0	202.1	172.3	156 4	224.6	128.4	119.4
15	208 2	252.9	202.1	171.5	296.1	149.9	170.1
16	245 2	334 6	233 3	193.5	344 3	188.6	151.6
17	207.6	334.5	193.4	163.0	274.3	217.7	144.4
18	104 0	175.2	101.0	96.7	125.5	140.2	103 0
19	118.4	200.9	115.1	110.7	143.8	167.4	112 0
20	142 2	263.3	137.2	129 8	185.6	217.5	124.1
21	170 9	335.4	163 5	152 4	234.0	277.8	131 4
22	251 3	439.5	239 2	223.5	325.3	332 6	156 0
23	380.5	553.5	363 2	349.8	443.2	386 0	186.1
24	454.5	603.6	435 3	421.5	521.2	434 6	199 8

Table 3 (cont)

2/22/2001

Group Quarters Population Sex Ratio by Single Year of Age and Race (IMPRACE)

Sex Ratios - 2000

			Not Hispanic				
Age	Total POP	Hispanic	Total	White	Black or African American	American Indian and Alaska Native	Asian and Pacific Islander
25	491.9	716 0	459.0	390.9	695.3	359.5	196 6
26	532 4	759.7	498.3	420.0	734.6	409.7	207.6
27	556 2	768 0	522.4	440.0	767.9	366.0	211.4
28	548 0	718 2	521.0	435.4	742.2	412.3	217 1
29	536.3	734.5	506.2	424.3	700.6	367.2	233 4
30	522 9	758.6	489.7	415.2	661.5	371.0	225.9
31	510 5	740.8	478.3	400 9	636.1	368 8	256.2
32	494.7	713.4	464.4	396.3	605.7	352.5	223.6
33	503 5	690 8	477 2	410.3	603 0	373.2	264 0
34	489.6	691.7	463 0	396.6	583 8	350 3	273 9
35	466.8	659.3	442.5	388.2	545.5	342.7	255 2
36	468 0	633 2	447 5	392.1	552.6	381 2	269.9
37	471.5	678 6	448 2	395.1	552.2	330.6	269.3
38	459.3	651 4	438 5	385 9	547.2	332.2	247.0
39	453 0	642 6	432.9	378.1	548 9	337.5	226.5
40	444.4	601.4	427 2	372.3	535 2	382.7	249 5
41	441 8	605 9	424.7	363 5	549.5	393.3	243 8
42	430 1	633 4	410 7	345.3	552 0	368 5	237.6
43	426.3	613 0	408 0	342 0	558.3	347 4	221.1
44	414.2	670 7	392.0	324.5	553 2	309 8	221.6
45	394 7	606 8	376.3	316 3	527 2	324 3	197.1
46	385 8	600.5	367 1	301.0	550.1	338 4	191.2
47	363 6	553 1	347.3	290 2	517 0	320 0	175 5
48	358 4	532.3	343 5	285.8	528.1	337 9	175 8
49	345 6	523 5	331 0	269.6	521 8	363 0	208.0
50	316 0	513 2	300 9	251.0	476.6	288 5	164.8
51	301 9	453 0	290 2	241.9	467.0	327.2	156.9
52	271 6	415.2	261 6	222.1	416.7	328 5	167.7

Sex Ratios - 1990

			Not Hispanic				
Age	Total POP	Hispanic	Total	White	Black or African American	American Indian and Alaska Native	Asian and Pacific Islander
25	471.1	618.2	451.3	428.0	543.4	443.3	214.9
26	491.0	654.6	468.7	439.0	574.5	375.7	222.0
27	493 7	678.2	469.0	431.2	564.1	463.1	267 5
28	497 8	694.9	472.2	425.2	587.4	397.9	261.1
29	497.1	672.2	472.3	421.7	592.4	391.7	245 5
30	490.8	654 2	468.4	416.7	583.9	433.5	237.2
31	493.8	679 2	469.8	405.9	614.7	458 4	227.2
32	489.7	684.0	465.4	400.6	609.3	426 1	253.3
33	479.0	673.2	455.1	389.2	602.2	432 6	235 9
34	473.0	660.3	449.3	383.6	599.3	403.9	241 3
35	463.9	670 2	439.6	366.3	613.5	440.4	244.0
36	461.9	676.1	438.2	368.0	621.8	391.7	240.2
37	455.5	629.4	435.2	366.1	613.3	380.4	282.6
38	450.6	608.1	432.6	357.8	654.8	390.2	237.6
39	442.3	652 4	419.0	342.4	647.9	433 8	231.3
40	423 5	627 6	402.3	330.2	629.9	482.9	229 7
41	404.5	630.1	383.4	318.8	612.7	429.1	201 6
42	376.8	637.7	355.0	294.1	613.6	337.4	209.0
43	351.8	624.5	329.5	277.5	659.7	372 5	200.6
44	336.3	641.1	311.2	257 1	550 5	407 3	227 8
45	313.5	574 5	292.8	246.2	499.7	413 2	206 6
46	287.0	525 3	269.9	229.6	481.3	364 8	188 0
47	269.5	568.7	249.8	212.5	444.2	343 6	178 7
48	258 3	504 5	241.4	205.2	439.1	362 7	177 9
49	258.8	515.6	240.2	202 6	442.3	382 4	196 8
50	247 4	536.0	229.3	197.9	404.7	272 0	185 7
51	219.9	502.7	204.4	179.3	344.2	337 9	149.5
52	208.1	441.5	195.1	170.6	345.2	242.9	136.7

Table 3 (cont)

2/22/2001

Group Quarters Population Sex Ratio by Single Year of Age and Race (IMPRACE)

	Sex Ratios - 2000		Not Hispanic					Sex Ratios - 1990		Not Hispanic				
Age	Total POP	Hispanic	Total	White	Black or African American	American Indian and Alaska Native	Asian and Pacific Islander	Total POP	Hispanic	Total	White	Black or African American	American Indian and Alaska Native	Asian and Pacific Islander
53	259.2	489.2	245.6	206.3	436.0	304.5	130.3	206.4	448.1	192.4	167.8	338.0	532.1	144.2
54	240.9	399.3	229.5	193.7	405.1	266.7	116.7	194.6	443.8	181.3	161.6	295.6	253.7	143.6
55	234.0	410.3	222.8	193.3	366.1	247.7	132.4	187.3	420.8	175.6	156.7	279.1	313.0	142.0
56	204.7	386.4	194.8	171.1	318.6	275.5	115.1	173.5	367.4	164.9	148.3	271.2	263.4	152.9
57	189.5	348.3	180.7	157.2	305.4	253.6	131.1	163.3	349.2	155.0	139.2	261.8	250.0	114.0
58	184.4	309.7	176.6	157.0	278.6	258.8	105.2	155.5	319.3	148.5	136.5	232.8	313.9	108.1
59	178.9	307.3	171.4	151.4	276.9	254.4	118.1	151.8	297.3	144.9	132.1	229.7	374.3	127.7
60	169.3	289.3	162.4	145.5	245.6	195.2	113.1	142.2	301.7	135.6	125.4	204.0	208.1	118.4
61	160.6	293.0	153.7	138.4	232.1	204.9	122.7	136.1	254.4	131.6	122.3	200.4	229.3	107.9
62	146.3	242.1	140.9	128.1	202.6	205.1	141.5	126.0	216.8	122.7	114.9	175.5	161.3	119.5
63	137.1	203.2	133.3	120.9	192.7	291.7	104.2	119.3	201.3	116.3	108.7	169.5	259.0	108.9
64	129.9	206.9	125.7	116.0	181.2	178.0	91.7	111.4	198.2	108.5	102.5	149.0	196.3	98.1
65	122.2	204.3	118.2	108.9	162.9	191.5	105.5	106.8	188.0	104.4	99.0	143.8	162.0	99.4
66	113.9	193.5	110.2	101.7	158.5	159.6	84.6	98.7	160.4	96.9	92.2	132.3	130.5	93.0
67	105.0	170.7	102.0	94.3	143.6	121.4	85.3	93.6	145.4	92.1	87.6	122.3	220.5	109.4
68	97.0	138.8	95.2	88.7	134.7	120.6	76.8	87.7	137.0	86.4	82.2	116.9	132.8	120.5
69	91.6	126.9	90.0	83.9	124.3	151.2	97.7	82.5	135.3	81.1	76.8	111.3	164.3	116.7
70	87.3	126.0	85.6	80.8	112.4	110.4	88.9	77.7	127.3	76.6	73.4	100.8	116.9	92.0
71	81.6	118.4	80.1	75.5	105.9	133.7	94.7	71.6	101.1	71.1	68.5	91.2	128.6	93.0
72	75.4	114.3	74.0	70.3	97.5	93.9	78.2	66.3	96.2	65.7	63.6	81.5	119.4	89.8
73	68.7	103.4	67.5	64.0	89.8	99.1	73.5	61.4	87.4	60.9	59.3	74.3	90.6	77.2
74	64.6	96.4	63.5	61.1	78.9	78.3	74.9	57.3	92.8	56.6	54.7	73.2	68.2	91.7
75	60.2	87.6	59.4	56.2	80.6	101.6	73.9	54.0	74.8	53.6	51.7	70.8	87.5	74.4
76	57.0	89.9	56.1	53.8	72.4	83.3	76.7	50.2	74.9	49.6	48.2	64.5	79.1	76.4
77	52.2	72.4	51.7	49.3	67.7	83.1	81.1	46.2	67.3	45.8	44.1	61.1	100.0	86.6
78	48.8	71.9	48.2	46.5	62.3	80.8	63.7	43.2	60.4	42.9	41.5	58.8	76.7	68.2
79	44.8	65.1	44.3	42.8	59.5	59.7	54.9	40.1	58.7	39.8	38.4	53.5	73.8	78.9
80	42.7	66.8	42.2	40.4	56.6	51.4	66.3	37.8	59.6	37.4	36.1	52.1	57.4	71.0

Table (3)

2/22/2001

Group Quarters Population Sex Ratio by Single Year of Age and Race (IMPRACE)

Sex Ratios - 2000

Age	Total POP	Hispanic	Not Hispanic: Total	Not Hispanic: White	Not Hispanic: Black or African American	Not Hispanic: American Indian and Alaska Native	Not Hispanic: Asian and Pacific Islander
81	39.9	57.9	39.5	38.3	50.3	67.1	56 3
82	37.5	59.8	37.0	36.4	43 1	56.8	48.6
83	34.7	46.1	34.5	33.6	42.1	51.4	58.5
84	32.7	47.3	32.4	31.8	38 7	44 5	46 6
85	31 0	40.3	30.8	30 3	36.3	47 6	43 1
86	29 2	43 4	28 9	28 5	32 6	50.9	47 5
87	28 0	38 9	27.8	27 2	32.2	51.4	46.4
88	25.9	34 2	25 7	25.3	30.1	36.7	43 5
89	24 8	35.5	24.6	24.1	28.6	39 3	46.4
90	23.2	39 5	22.9	22 4	27 9	37.0	40 1
91	21.3	30 8	21.1	20 7	25 5	33.3	45 4
92	20 7	34 0	20.5	20.0	24.8	36 5	45 2
93	19 2	34.2	19.0	18.7	21.7	27 5	38 7
94	18 3	33 4	18.1	17.7	21.7	47 2	36 7
95	17 4	29.0	17.2	16.7	22 6	20 6	34.5
96	16 5	33.4	16.2	15.7	20 0	52 6	35.9
97	15 3	29.7	15 1	14.6	20.0	35 5	28 6
98	15 0	31 8	14.7	14.2	20 7	21.9	34 0
99	15.2	32 8	15 0	14.2	23 2	26 3	26 5
100							

Sex Ratios - 1990

Age	Total POP	Hispanic	Not Hispanic: Total	Not Hispanic: White	Not Hispanic: Black or African American	Not Hispanic: American Indian and Alaska Native	Not Hispanic: Asian and Pacific Islander
81	34.6	53.1	34.3	33.3	47.3	59.1	68.7
82	32.7	52.9	32.4	31.4	45.7	68 0	64.9
83	30.8	52.9	30.5	29.7	41.5	53.5	71.0
84	29.1	47.0	28.9	28.0	39 6	54 0	61.2
85	27 5	45.6	27.2	26.4	38.3	74.3	53.8
86	25 9	46.1	25.6	24.9	35.5	52 6	53 4
87	25.1	44.2	24.8	24.1	36.3	54 8	46 5
88	24.2	46.6	23.9	23.4	32.5	46.7	40 2
89	23.1	43.3	22 9	22.2	32.1	52.6	41.3
90	22.5	40.8	22.3	21.8	29.2	58 5	43.7
91	21.6	44.3	21.4	20.9	30 4	41.0	40.4
92	20.5	35.3	20.3	20.1	23.7	44 6	31.0
93	20.0	38 9	19.8	19.5	25.7	33 3	28 5
94	19.4	36 5	19.2	18.9	25.1	31 3	25.7
95	19.3	40.0	19.1	18.8	23.4	34.2	21.4
96	18.5	37.4	18.3	18 1	21.9	45 5	21.2
97	18.0	34.6	17 8	17.5	24 5	38 9	23.1
98	16 4	24.2	16 3	16.0	19.5	11 1	38 8
99	17.4	30.6	17.2	16.7	22.9	36 4	27 6
100	16.5	31 1	16 3	15.6	22.6	31 4	23 4

filename - sexratiogq20001990

Table 4

Sex Ratios by Age and Sex: CPS 2000					
Age	Hispanic	White non-Hispanic	Black non-Hispanic	AIAN non-Hispanic	API non-Hispanic
Under 5 years	103	107	99	67	104
5 to 19 years	106	105	103	109	104
20 to 24 years	112	100	85	97	106
25 to 29 years	100	100	77	67	93
30 to 34 years	103	99	83	78	100
35 to 44 years	99	101	85	93	82
45 to 54 years	97	97	85	79	90
55 to 64 years	89	94	73	114	85
65 to 74 years	71	85	69	69	86
75 to 84 years	67	69	67	49	107
85 years and over	71	49	48	132	62

filename: sex ration by age sex caps cps march00

Table 5

2/22/01

Allocation Rates1/ by Age, Sex, Race and Hispanic Origin: 2000 and 1990

	Total Population 2000	Total Population 1990	Group Quarters 2000	Group Quarters 1990
Age	3.7	2.4	7.6	4.2
Sex	1.0	1.2	3.1	1.4
Race	4.0	2.0	9.1	4.3
Hispanic	2.4	10.0	12.3	12.9

1/ Excludes substitution

filename: allocation age sex race hispanic origin

Table 6

INFORMATION DELETED

ESCAP MEETING NO. 45 - 02/23/01

MINUTES

Minutes of the Executive Steering Committee on Accuracy and Coverage Evaluation (A.C.E.) Policy (ESCAP) Meeting # 45

February 23, 2001

Prepared by: Sarah Brady

The forty-fifth meeting of the Executive Steering Committee on Accuracy and Coverage Evaluation Policy was held on February 23, 2001 at 1:30. The agenda for the meeting was to present results from the sensitivity analysis of the loss functions and to present sex ratios for group quarters.

Committee Attendees:

Nancy Potok
Cynthia Clark
John Thompson
Jay Waite
Bob Fay
Howard Hogan
Ruth Ann Killion
John Long
Carol Van Horn

Deputy Director/Acting Director:
William Barron

Other Attendees:

Marvin Raines
William Bell
Donna Kostanich
Alfredo Navarro
Maria Urrutia
Sarah Brady

I. Loss Functions

Alfredo Navarro presented various loss function simulations. The handouts from the presentation are attached. The loss functions presented illustrated sensitivity analyses where levels of processing error and correlation bias assumptions (various models and levels) were varied. DSSD staff are currently in the process of generating loss functions for counties. The Committee noted that correlation bias has a significant effect on loss functions. The loss functions indicated improvement when full correlation bias is assumed regardless of the model used. Moreover, even with 50 percent correlation bias an improvement is noted.

II. Sex Ratios for Group Quarters

John Long presented data on sex ratios and other characters for group quarters (GQs) in 1990 and 2000. The Committee noted that there were not any unusual results. Thus, it was concluded that the difference between A.C.E. and DA were not due to the way GQs were enumerated in the census.

III. Miscellaneous Items

John Long and Howard Hogan updated the Committee on their research to explain the difference between A.C.E. and DA by reevaluating the PES and DA results from 1990. They will continue their work over the weekend and will present their findings to the Committee on Monday.

Bob Fay discussed some preliminary findings from his research on TES and the potential balancing error. He will present his preliminary report on Monday.

The Committee then held a private deliberation session chaired by John Thompson. Concerns were expressed that there is limited time left for them to come to a recommendation and they must reach a conclusion by close of business Monday.

IV. Next Meeting

The agenda for the next meeting, scheduled for February 26, 2001, is to discuss remaining issues with DA and loss functions for counties and to examine the loss functions results correcting for synthetic bias.

ESCAP MEETING NO. 46 - 02/26/01

AGENDA

There was no agenda developed or used for the February 26, 2001 meeting.

ESCAP MEETING NO. 46 - 02/26/01

HANDOUTS

Materials attached to these minutes were draft and preliminary material to inform the ESCAP Committee. The data and analysis contained in these documents are subject to revision and are not final. These materials report the results of research and analysis undertaken by Census Bureau staff. They have undergone a more limited review than official Census Bureau publications. Research results and conclusions expressed are those of the authors and do not necessarily indicate concurrence by the Census Bureau.

Places By Total Unadjusted Population and Adjusted Population Difference
(Excludes New Mexico)

TABLE 1 OF UNADJ BY DIFF
CONTROLLING FOR SUMLEV=050 (County)

UNADJ(Unadjusted Population) DIFF(Diference)

Frequency Percent Row Pct Col Pct	-100+	-50-99	-10-49	-5-9	-1-4	0	1-4	5-9	10-49	50-99	100+	Total
1-24	0 0.00 . 0.00	0 0.00 . 0.00	0 0.00 . 0.00	0 0.00 . 0.00	0 0.00 . 0.00	0 0.00 . 0.00	0 0.00 . 0.00	0 0.00 . 0.00	0 0.00 . 0.00	0 0.00 . 0.00	0 0.00 . 0.00	0 0.00
25-49	0 0.00 . 0.00	0 0.00 . 0.00	0 0.00 . 0.00	0 0.00 . 0.00	0 0.00 . 0.00	0 0.00 . 0.00	0 0.00 . 0.00	0 0.00 . 0.00	0 0.00 . 0.00	0 0.00 . 0.00	0 0.00 . 0.00	0 0.00
50-99	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	1 0.03 100.00 6.25	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	1 0.03
100-199	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	1 0.03 100.00 4.17	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	1 0.03
200-499	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	2 0.06 50.00 8.33	1 0.03 25.00 2.86	1 0.03 25.00 0.42	0 0.00 0.00 0.00	0 0.00 0.00 0.00	4 0.13
500-999	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	2 0.06 8.33 5.00	6 0.19 25.00 16.67	4 0.13 16.67 25.00	1 0.03 4.17 4.17	3 0.10 12.50 8.57	8 0.26 33.33 3.36	0 0.00 0.00 0.00	0 0.00 0.00 0.00	24 0.77
1,000+	36 1.16 1.17 100.00	60 1.93 1.95 100.00	258 8.30 8.38 100.00	38 1.22 1.23 95.00	30 0.97 0.97 83.33	11 0.35 0.36 68.75	20 0.64 0.65 83.33	31 1.00 1.01 88.57	229 7.37 7.44 96.22	317 10.20 10.30 100.00	2048 65.89 66.54 100.00	3078 99.03
Total	36 1.16	60 1.93	258 8.30	40 1.29	36 1.16	16 0.51	24 0.77	35 1.13	238 7.66	317 10.20	2048 65.89	3108 100.00

Places By Total Unadjusted Population and Adjusted Population Difference

(Excludes New Mexico)

TABLE 2 OF UNADJ BY DIFF

CONTROLLING FOR SUMLEV=060 (County Subdivision)

UNADJ(Unadjusted Population) DIFF(Diference)

Frequency Percent Row Pct Col Pct	-100+	-50-99	-10-49	-5-9	-1-4	0	1-4	5-9	10-49	50-99	100+	Total
1-24	0	0	0	0	3	498	2	0	0	0	0	503
	0.00	0.00	0.00	0.00	0.01	1.42	0.01	0.00	0.00	0.00	0.00	1.43
	0.00	0.00	0.00	0.00	0.60	99.01	0.40	0.00	0.00	0.00	0.00	
	0.00	0.00	0.00	0.00	0.07	7.59	0.06	0.00	0.00	0.00	0.00	
25-49	0	0	0	0	37	943	17	0	0	0	0	997
	0.00	0.00	0.00	0.00	0.11	2.69	0.05	0.00	0.00	0.00	0.00	2.84
	0.00	0.00	0.00	0.00	3.71	94.58	1.71	0.00	0.00	0.00	0.00	
	0.00	0.00	0.00	0.00	0.85	14.38	0.48	0.00	0.00	0.00	0.00	
50-99	0	0	0	0	182	1299	95	1	0	0	0	1577
	0.00	0.00	0.00	0.00	0.52	3.70	0.27	0.00	0.00	0.00	0.00	4.49
	0.00	0.00	0.00	0.00	11.54	82.37	6.02	0.06	0.00	0.00	0.00	
	0.00	0.00	0.00	0.00	4.17	19.80	2.69	0.04	0.00	0.00	0.00	
100-199	0	0	0	3	490	1400	301	11	2	0	0	2207
	0.00	0.00	0.00	0.01	1.40	3.99	0.86	0.03	0.01	0.00	0.00	6.29
	0.00	0.00	0.00	0.14	22.20	63.43	13.64	0.50	0.09	0.00	0.00	
	0.00	0.00	0.00	0.18	11.22	21.34	8.53	0.44	0.03	0.00	0.00	
200-499	0	0	3	139	1794	1587	951	389	63	0	0	4926
	0.00	0.00	0.01	0.40	5.11	4.52	2.71	1.11	0.18	0.00	0.00	14.03
	0.00	0.00	0.06	2.82	36.42	32.22	19.31	7.90	1.28	0.00	0.00	
	0.00	0.00	0.25	8.22	41.07	24.19	26.95	15.72	0.80	0.00	0.00	
500-999	0	0	92	792	1140	446	981	596	807	0	0	4854
	0.00	0.00	0.26	2.26	3.25	1.27	2.79	1.70	2.30	0.00	0.00	13.83
	0.00	0.00	1.90	16.32	23.49	9.19	20.21	12.28	16.63	0.00	0.00	
	0.00	0.00	7.61	46.84	26.10	6.80	27.80	24.09	10.19	0.00	0.00	
1,000+	65	63	1114	757	722	387	1182	1477	7048	2918	4310	20043
	0.19	0.18	3.17	2.16	2.06	1.10	3.37	4.21	20.08	8.31	12.28	57.09
	0.32	0.31	5.56	3.78	3.60	1.93	5.90	7.37	35.16	14.56	21.50	
	100.00	100.00	92.14	44.77	16.53	5.90	33.49	59.70	88.99	100.00	100.00	
Total	65	63	1209	1691	4368	6560	3529	2474	7920	2918	4310	35107
	0.19	0.18	3.44	4.82	12.44	18.69	10.05	7.05	22.56	8.31	12.28	100.00

Places By Total Unadjusted Population and Adjusted Population Difference

(Excludes New Mexico)

TABLE 3 OF UNADJ BY DIFF

CONTROLLING FOR SUMLEV=160 (PLACE)

UNADJ(Unadjusted Population) DIFF(Diference)

Frequency Percent Row Pct Col Pct	-100+	-50-99	-10-49	-5-9	-1-4	0	1-4	5-9	10-49	50-99	100+	Total
1-24	0	0	0	0	0	157	4	0	0	0	0	161
	0.00	0.00	0.00	0.00	0.00	0.63	0.02	0.00	0.00	0.00	0.00	0.65
	0.00	0.00	0.00	0.00	0.00	97.52	2.48	0.00	0.00	0.00	0.00	
	0.00	0.00	0.00	0.00	0.00	5.58	0.11	0.00	0.00	0.00	0.00	
25-49	0	0	0	0	8	237	32	0	0	0	0	277
	0.00	0.00	0.00	0.00	0.03	0.95	0.13	0.00	0.00	0.00	0.00	1.11
	0.00	0.00	0.00	0.00	2.89	85.56	11.55	0.00	0.00	0.00	0.00	
	0.00	0.00	0.00	0.00	0.33	8.42	0.92	0.00	0.00	0.00	0.00	
50-99	0	0	0	0	113	569	175	2	0	0	0	859
	0.00	0.00	0.00	0.00	0.45	2.29	0.70	0.01	0.00	0.00	0.00	3.45
	0.00	0.00	0.00	0.00	13.15	66.24	20.37	0.23	0.00	0.00	0.00	
	0.00	0.00	0.00	0.00	4.69	20.22	5.02	0.08	0.00	0.00	0.00	
100-199	0	0	0	3	381	749	639	36	1	0	0	1809
	0.00	0.00	0.00	0.01	1.53	3.01	2.57	0.14	0.00	0.00	0.00	7.27
	0.00	0.00	0.00	0.17	21.06	41.40	35.32	1.99	0.06	0.00	0.00	
	0.00	0.00	0.00	0.34	15.81	26.62	18.34	1.51	0.02	0.00	0.00	
200-499	0	0	7	140	956	684	1400	710	180	0	0	4077
	0.00	0.00	0.03	0.56	3.84	2.75	5.62	2.85	0.72	0.00	0.00	16.37
	0.00	0.00	0.17	3.43	23.45	16.78	34.34	17.41	4.42	0.00	0.00	
	0.00	0.00	1.25	15.80	39.67	24.31	40.18	29.74	2.78	0.00	0.00	
500-999	0	0	91	384	539	235	624	766	1071	3	0	3713
	0.00	0.00	0.37	1.54	2.16	0.94	2.51	3.08	4.30	0.01	0.00	14.91
	0.00	0.00	2.45	10.34	14.52	6.33	16.81	20.63	28.84	0.08	0.00	
	0.00	0.00	16.28	43.34	22.37	8.35	17.91	32.09	16.51	0.14	0.00	
1,000+	31	58	461	359	413	183	610	873	5234	2173	3608	14003
	0.12	0.23	1.85	1.44	1.66	0.73	2.45	3.51	21.02	8.73	14.49	56.24
	0.22	0.41	3.29	2.56	2.95	1.31	4.36	6.23	37.38	15.52	25.77	
	100.00	100.00	82.47	40.52	17.14	6.50	17.51	36.57	80.70	99.86	100.00	
Total	31	58	559	886	2410	2814	3484	2387	6486	2176	3608	24899
	0.12	0.23	2.25	3.56	9.68	11.30	13.99	9.59	26.05	8.74	14.49	100.00

Frequency Missing = 7

Places By Total Unadjusted Population and Adjusted Population Difference

(Excludes New Mexico)

TABLE 4 OF UNADJ BY DIFF
CONTROLLING FOR SUMLEV=280 (AIR)

UNADJ(Unadjusted Population) DIFF(Diference)

Frequency Percent Row Pct Col Pct	-100+	-50-99	-10-49	-5-9	-1-4	0	1-4	5-9	10-49	50-99	100+	Total
1-24	0	0	0	0	0	52	4	0	0	0	0	56
	0.00	0.00	0.00	0.00	0.00	8.75	0.67	0.00	0.00	0.00	0.00	9.43
	0.00	0.00	0.00	0.00	0.00	92.86	7.14	0.00	0.00	0.00	0.00	
	.	.	.	.	0.00	23.53	5.41	0.00	0.00	0.00	0.00	
25-49	0	0	0	0	0	14	16	0	0	0	0	30
	0.00	0.00	0.00	0.00	0.00	2.36	2.69	0.00	0.00	0.00	0.00	5.05
	0.00	0.00	0.00	0.00	0.00	46.67	53.33	0.00	0.00	0.00	0.00	
	.	.	.	.	0.00	6.33	21.62	0.00	0.00	0.00	0.00	
50-99	0	0	0	0	0	25	36	3	0	0	0	64
	0.00	0.00	0.00	0.00	0.00	4.21	6.06	0.51	0.00	0.00	0.00	10.77
	0.00	0.00	0.00	0.00	0.00	39.06	56.25	4.69	0.00	0.00	0.00	
	.	.	.	.	0.00	11.31	48.65	7.14	0.00	0.00	0.00	
100-199	0	0	0	0	2	43	13	29	6	0	0	93
	0.00	0.00	0.00	0.00	0.34	7.24	2.19	4.88	1.01	0.00	0.00	15.66
	0.00	0.00	0.00	0.00	2.15	46.24	13.98	31.18	6.45	0.00	0.00	
	.	.	.	.	66.67	19.46	17.57	69.05	4.88	0.00	0.00	
200-499	0	0	0	0	0	57	3	9	43	0	0	112
	0.00	0.00	0.00	0.00	0.00	9.60	0.51	1.52	7.24	0.00	0.00	18.86
	0.00	0.00	0.00	0.00	0.00	50.89	2.68	8.04	38.39	0.00	0.00	
	.	.	.	.	0.00	25.79	4.05	21.43	34.96	0.00	0.00	
500-999	0	0	0	0	0	29	1	1	47	0	0	78
	0.00	0.00	0.00	0.00	0.00	4.88	0.17	0.17	7.91	0.00	0.00	13.13
	0.00	0.00	0.00	0.00	0.00	37.18	1.28	1.28	60.26	0.00	0.00	
	.	.	.	.	0.00	13.12	1.35	2.38	38.21	0.00	0.00	
1,000+	0	0	0	0	1	1	1	0	27	32	99	161
	0.00	0.00	0.00	0.00	0.17	0.17	0.17	0.00	4.55	5.39	16.67	27.10
	0.00	0.00	0.00	0.00	0.62	0.62	0.62	0.00	16.77	19.88	61.49	
	.	.	.	.	33.33	0.45	1.35	0.00	21.95	100.00	100.00	
Total	0	0	0	0	3	221	74	42	123	32	99	594
	0.00	0.00	0.00	0.00	0.51	37.21	12.46	7.07	20.71	5.39	16.67	100.00

Materials attached to these minutes were draft and preliminary material to inform the ESCAP Committee. The data and analysis contained in these documents are subject to revision and are not final. These materials report the results of research and analysis undertaken by Census Bureau staff. They have undergone a more limited review than official Census Bureau publications. Research results and conclusions expressed are those of the authors and do not necessarily indicate concurrence by the Census Bureau.

DSSD Census 2000 Procedures and Operations Memorandum Series B-4
DRAFT - February 23, 2001

Accuracy and Coverage Evaluation: Demographic Analysis Results

[EXECUTIVE SUMMARY ONLY]

J. Gregory Robinson
U.S. Census Bureau

Accuracy and Coverage Evaluation 2000: Demographic Analysis Results

Prepared by J Gregory Robinson

Executive Summary

We use Demographic Analysis population estimates to:

1) assess the completeness of coverage in Census 2000 and document the change in coverage from previous censuses,

2) to check the consistency of the survey-based A.C.E. coverage estimates with the DA estimates.

What was the magnitude of net undercount in Census 2000 as measured by DA and the A.C.E.? Do they agree?

The A.C.E. measures a net undercount of 3.3 million, or 1.2 percent for Census 2000.

DA measures a lower net undercount than the A.C.E., according to either of the two sets of DA estimates developed. The "base" DA set estimates a net overcount of 1.8 million, or 0.7 percent in 2000. The "alternative" set, which increases the DA estimate to allow for additional undocumented immigration in the 1990's, gives a net undercount of 0.9 million, or 0.3 percent.

What do DA and A.C.E. say about change in net undercount from 1990?

The DA and A.C.E. estimates both measure a reduction in net undercount in Census 2000 compared to 1990, but DA implies a greater change. Under the base set, the estimated DA net undercount rate fell by 2.5 percentage points from 1.8 percent net undercount in 1990 to 0.7 percent net overcount in 2000. Under the alternative DA set, the net undercount rate was reduced by 1.5 percentage points from 1.8 percent in 1990 to 0.3 percent net undercount in 2000.

The A.C.E. estimate of 1.2 percent net undercount in 2000 was 0.4 percentage points lower than the 1.6 percent in 1990.

Do DA and A.C.E. measure reductions in the differential undercount?

Yes, both DA and A.C.E measure a reduction in the net undercount rates of Black and Nonblack children (ages 0-17) compared to 1990. Both methods also measure a reduction in the net undercount rates of Black men and women (ages 18+).

Where do the DA and A.C.E. estimates disagree?

DA finds a reduction in the net undercount rates of Nonblack men and women in Census 2000 compared to the rates of previous censuses. The reduction is large under the base DA set and moderate under the alternative DA set.

The A C E. indicates no change or a slight increase in undercount rates for Nonblack adults as a group.

Has "correlation bias" in the survey estimates been reduced?

No. The A.C.E. sex ratios (ratio of males per 100 females) for Black adults are much lower than DA "expected"sex ratios, implying that A.C.E. is not capturing the high undercount rates of Black men relative to Black women (the well-known "correlation bias"). The size of this bias is about the same as in the 1990 Post Enumeration Survey

Do other demographic benchmarks support the DA finding of a large reduction in net undercount in Census 2000 compared to 1990 and previous censuses?

Yes. the comparison of census counts to auxiliary data sets (such as school enrollment data for children and Medicare enrollment for the population 65 and older) are consistent in indicating Census 2000 is more complete relative to 1990.

Is Alternative DA Reasonable?

Although the alternative DA set may have gone too far in doubling the flow of undocumented immigrants during the decade, it does provide a reasonable benchmark. The data indicate that we may have understated the immigration component in the base DA. However, increasing the flow of undocumented immigrants during the decade to reach the ACE totals results in percent Hispanic and percent foreign born that are even higher relative to CPS 2000.

Table 9A. Percent Hispanic

	% Hispanic
1990 Census	8.99
Census 2000	12.55
Implied by controlling to A.C.E.	13.20
Implied by Base DA	12.13
Implied by Alternate DA	12.72

[1] The undocumented population is increased by the amount necessary to reach the A.C.E. population estimate.

Table 9B. Percent Foreign Born

	Total	Hispanic	Non Hispanic
1990 Census	7.95	35.81	5.26
Reweighted CPS 2000[1]	10.61	39.14	6.44
Implied by controlling to A.C.E.[2]	11.89	42.94	7.16
Implied by Base DA	10.26	36.52	6.63
Implied by Alternate DA	11.13	40.05	6.92

[1] The CPS 2000 figure is weighted to the Census 2000

[2] The undocumented population is increased by the amount necessary to reach .the A.C.E. population estimate

Materials attached to these minutes were draft and preliminary material to inform the ESCAP Committee. The data and analysis contained in these documents are subject to revision and are not final. These materials report the results of research and analysis undertaken by Census Bureau staff. They have undergone a more limited review than official Census Bureau publications. Research results and conclusions expressed are those of the authors and do not necessarily indicate concurrence by the Census Bureau.

DSSD Census 2000 Procedures and Operations Memorandum Series B-4
DRAFT - February 23, 2001

Accuracy and Coverage Evaluation: Demographic Analysis Results

[Executive Summary Tables]

J. Gregory Robinson
U.S. Census Bureau

Revised 2-21-01

Table 10- Census Count, Demographic Analysis (DA) Estimates, and Accuracy and Coverage Evaluation (A.C.E.) Estimate for the U.S. Resident Population: 4-1-2000

	Count or Estimate
1. Census Count	281,421,906
2. D.A. Estimate	
a. Base Set	279,598,121
b. Alternative Set	282,335,711
3. A.C.E. Estimate	284,683,785
Difference from Census:	
4. D.A. Estimate	
a. Base Set (=2a-1)	(1,823,785)
b. Alternative Set (=2b-1)	913,805
5. A.C.E. Estimate (=3-1)	3,261,879
Percent Difference	
4. D.A. Estimate	
a. Base Set (=4a/2a*100)	-0.65
b. Alternative Set (=4b/2b*100)	0.32
5. A.C.E. Estimate (=5/3*100)	1.15

Note: The DA estimates for ages under 65 are based on components of population change (births, deaths, legal immigration, and estimates of emigration and undocumented immigration). The DA estimates for ages 65 and over are based on 2000 Medicare data, adjusted for underenrollment.

The A.C.E. and DA estimates are preliminary.

D.A. Base Set - DA estimates without alternative assumptions.
D.A. Alternative Set - DA estimates with alternative assumption that doubles the estimated number of undocumented immigrants entering during the 1990's (from 2.75 to 5.5 million).

Revised 2-21-01

Table 12--Estimates of Percent Net Undercount by Sex and Age: 1960 to 2000
(a minus sign denotes a net overcount)

	Demographic Analysis						Survey-based	
					2000		PES	A.C.E.
Category	1960	1970	1980	1990	Base DA	Alt DA	1990	2000
MALE								
Total	3.5	3.4	2.2	2.8	**-0.1**	**0.9**	1.9	**1.5**
0-17	2.8	2.7	0.9	2.2	**-0.5**	**0.3**	3.2	**1.5**
18-29	5.9	3.9	3.3	2.2	**-2.6**	**0.3**	3.2	**3.5**
30-49	4.2	5.1	3.6	3.8	**1.3**	**2.3**	1.9	**1.8**
50+	2.2	2.5	1.2	2.7	**0.2**	**0.3**	-0.6	**-0.2**
FEMALE								
Total	2.7	2.0	0.3	0.9	**-1.2**	**-0.2**	1.3	**0.8**
0-17	1.8	2.4	0.9	2.4	**0.1**	**0.9**	3.2	**1.5**
18-29	2.8	1.3	0.4	0.6	**-3.1**	**-0.7**	2.8	**2.1**
30-49	1.9	1.3	-0.0	0.5	**-0.9**	**0.0**	0.9	**1.0**
50+	4.6	2.6	-0.2	0.2	**-1.4**	**-1.3**	-1.2	**-0.8**

Note: DA estimates are consistent with estimates in Table 2.

Revised 2-24-01

Table 14--Estimates of Percent Net Undercount by Race, Sex and Age: 1960 to 2000
(a minus sign denotes a net overcount)

	Demographic Analysis										Survey-based	
					2000-Base DA			2000-Alt DA			PES	A.C.E.
Category	1960	1970	1980	1990	Average	Model 1	Model 2	Average	Model 1	Model 2	1990	2000
BLACK MALE												
Total	8.8	9.1	7.5	8.5	**5.1**	6.9	3.3	**5.8**	7.6	4.0	4.9	**2.4**
0-17	5.4	6.2	4.2	5.9	**1.5**	4.9	-1.9	**1.8**	5.1	-1.6	7.0	**3.0**
18-29	15.1	12.1	9.2	7.7	**6.5**	8.0	4.9	**8.1**	9.6	6.5	3.6	**3.7**
30-49	11.9	14.5	13.1	12.3	**9.2**	10.1	8.3	**10.1**	11.0	9.1	6.3	**2.6**
50+	6.6	6.3	4.6	8.3	**3.3**	4.1	2.5	**3.4**	4.2	2.6	-0.4	**-0.7**
BLACK FEMALE												
Total	4.4	4.0	1.7	3.0	**0.6**	2.5	-1.3	**1.3**	3.2	-0.6	4.0	**1.8**
0-17	4.0	5.6	3.9	5.9	**1.9**	5.4	-1.6	**2.2**	5.7	-1.2	7.1	**3.0**
18-29	5.4	4.5	2.4	2.9	**0.1**	1.9	-1.7	**1.8**	3.5	-0.0	5.5	**3.8**
30-49	2.1	0.5	0.6	2.5	**1.0**	2.1	-0.1	**1.9**	2.9	0.8	3.2	**1.3**
50+	7.6	3.8	-1.9	-0.8	**-1.3**	-0.5	-2.2	**-1.2**	-0.3	-2.0	-1.2	**-0.8**
NONBLACK MALE												
Total	2.9	2.7	1.5	2.0	**-0.9**	-1.2	-0.7	**0.2**	-0.1	0.4	1.5	**1.4**
0-17	2.4	2.1	0.3	1.5	**-0.9**	-1.6	-0.2	**-0.0**	-0.7	0.6	2.5	**1.3**
18-29	4.6	2.8	2.4	1.3	**-4.2**	-4.5	-3.9	**-1.0**	-1.3	-0.7	3.1	**3.4**
30-49	3.4	4.0	2.5	2.7	**-0.1**	-0.4	0.2	**1.1**	1.0	1.2	1.3	**1.7**
50+	1.8	2.2	0.9	2.2	**-0.2**	-0.2	-0.1	**-0.0**	-0.1	0.1	-0.6	**-0.2**
NONBLACK FEMALE												
Total	2.4	1.7	0.1	0.6	**-1.4**	-1.7	-1.1	**-0.5**	-0.8	-0.2	0.9	**0.6**
0-17	1.5	1.8	0.3	1.8	**-0.3**	-1.0	0.4	**0.6**	-0.1	1.3	2.5	**1.3**
18-29	2.4	0.9	0.1	0.3	**-3.7**	-4.0	-3.3	**-1.1**	-1.4	-0.8	2.4	**1.8**
30-49	1.9	1.3	-0.1	0.2	**-1.2**	-1.4	-1.0	**-0.2**	-0.4	-0.1	0.6	**0.9**
50+	4.3	2.5	-0.0	0.3	**-1.4**	-1.5	-1.4	**-1.3**	-1.4	-1.2	-1.2	**-0.8**

Sources and notes: See Table 2 and 4

Materials attached to these minutes were draft and preliminary material to inform the ESCAP Committee. The data and analysis contained in these documents are subject to revision and are not final. These materials report the results of research and analysis undertaken by Census Bureau staff. They have undergone a more limited review than official Census Bureau publications. Research results and conclusions expressed are those of the authors and do not necessarily indicate concurrence by the Census Bureau.

Table 14: Loss Function Synthetic Bias Correction for State Counts

Squared Error Loss				
Artificial Population	Census Loss minus A.C.E. Loss (1)	Synthetic Bias Correction (2)	Relative Bias (3)	Corrected Loss (4)
1	3.01E+11	9.16E+08	0.30%	3.02E+11
2	3.01E+11	5.58E+08	0.19%	3.01E+11
3	3.01E+11	9.09E+10	30.24%	3.92E+11
4	3.01E+11	-6.25E+10	-20.79%	2.38E+11

Table 15: Loss Function Synthetic Bias Correction for State Shares

Squared Error Loss				
Artificial Population	Census Loss minus A.C.E. Loss (1)	Synthetic Bias Correction (2)	Relative Bias (3)	Corrected Loss (4)
1	3.03E-07	-1.68E-08	-5.57%	2.86E-07
2	3.03E-07	-2.32E-11	-0.008%	3.03E-07
3	3.03E-07	4.08E-08	13.48%	3.43E-07
4	3.03E-07	-1.68E-07	-55.67%	1.34E-07

Table 16: Loss Function Synthetic Bias Correction for Congressional District Counts

Squared Error Loss				
Artificial Population	Census Loss minus A.C.E. Loss (1)	Synthetic Bias Correction (2)	Relative Bias (3)	Corrected Loss (4)
1	1.33E+10	-3.82E+08	-2.87%	1.29E+10
2	1.33E+10	-7.51E+07	-0.57%	1.32E+10
3	1.33E+10	3.97E+09	29.93%	1.73E+10
4	1.33E+10	-1.00E+09	-7.54%	1.23E+10

Table 17: Loss Function Synthetic Bias Correction for Congressional District Shares

Squared Error Loss				
Artificial Population	Census Loss minus A.C.E. Loss (1)	Synthetic Bias Correction (2)	Relative Bias (3)	Corrected Loss (4)
1	-4.27E-06	-3.24E-06	75.86%	-7.51E-06
2	-4.27E-06	-1.36E-06	31.79%	-5.63E-06
3	-4.27E-06	6.29E-05	-1470.99%	5.86E-05
4	-4.27E-06	2.30E-05	-538.88%	1.88E-05

Table 18: Equal CD Loss Function Synthetic Bias Correction for Congressional District Shares

Weighted Squared Error Loss (Weight = square of state census count)				
Artificial Population	Census Loss minus A.C.E. Loss (1)	Synthetic Bias Correction (2)	Relative Bias (3)	Corrected Loss (4)
1	1.61E+09	-2.14E+08	-13.27%	1.40E+09
2	1.61E+09	-4.60E+07	-2.85%	1.57E+09
3	1.61E+09	2.86E+09	177.17%	4.47E+09
4	1.61E+09	-4.84E+08	-29.98%	1.13E+09

Table 19: Weighted Loss Function Synthetic Bias Correction for State Counts

Weighted Squared Error Loss (Weight = 1/ census count)				
Artificial Population	Census Loss minus A.C.E. Loss (1)	Synthetic Bias Correction (2)	Relative Bias (3)	Corrected Loss (4)
1	1.79E+04	-126.95	-0.71%	1.77E+04
2	1.79E+04	-6.07	-0.03%	1.79E+04
3	1.79E+04	-1.60	-0.01%	1.79E+04
4	1.79E+04	-990.00	-5.54%	1.69E+04

Table 20: Weighted Loss Function Synthetic Bias Correction for State Shares

Weighted Squared Error Loss (Weight = 1/ census count)				
Artificial Population	Census Loss minus A.C.E. Loss (1)	Synthetic Bias Correction (2)	Relative Bias (3)	Corrected Loss (4)
1	5.92E-06	-4.38E-07	-7.40%	5.48E-06
2	5.92E-06	-2.09E-08	-0.35%	5.90E-06
3	5.92E-06	-5.53E-09	-0.09%	5.91E-06
4	5.92E-06	-3.41E-06	-57.69%	2.50E-06

Table 21: Weighted Loss Function Synthetic Bias Correction for Congressional District Shares

Weighted Squared Error Loss (Weight = 1/ census count)				
Artificial Population	Census Loss minus A.C.E. Loss (1)	Synthetic Bias Correction (2)	Relative Bias (3)	Corrected Loss (4)
1	2.07E+04	-4.99E+02	-2.41%	2.02E+04
2	2.07E+04	-8.69E+01	-0.42%	2.06E+04
3	2.07E+04	5.64E+03	27.22%	2.64E+04
4	2.07E+04	-1.61E+03	-7.79%	1.91E+04

Table 22: Weighted Loss Function Synthetic Bias Correction for Congressional District Shares

Weighted Squared Error Loss (Weight = 1/ census count)				
Artificial Population	Census Loss minus A.C.E. Loss (1)	Synthetic Bias Correction (2)	Relative Bias (3)	Corrected Loss (4)
1	2.09E-04	-2.51E-05	-12.04%	1.84E-04
2	2.09E-04	-7.73E-06	-3.70%	2.01E-04
3	2.09E-04	4.99E-04	238.79%	7.07E-04
4	2.09E-04	3.83E-05	18.36%	2.47E-04

Materials attached to these minutes were draft and preliminary material to inform the ESCAP Committee. The data and analysis contained in these documents are subject to revision and are not final. These materials report the results of research and analysis undertaken by Census Bureau staff. They have undergone a more limited review than official Census Bureau publications. Research results and conclusions expressed are those of the authors and do not necessarily indicate concurrence by the Census Bureau.

February 26, 2001

Additional Loss Function Analysis Results

What is the effect of correlation bias on the loss function results for counties?

Assumptions:

- Reduction of 10 percent in processing errors compared to 1990.
- Correlation bias range - 10% to 75%.

Even under the assumption that the **A.C.E.** realized modest gains in reducing processing errors (mostly matching error) the loss function analysis seems to be very robust with respect to the assumption of correlation bias.

For levels or numeric accuracy the results are more sensitive to correlation bias. (Table 1)

What is the effect of correlation bias on numeric accuracy for counties?

Assumptions:

- Correlation bias range - 10% to 75% for Blacks only.

The analysis was implemented for counties within size two categories:
Counties with over 100,000 population
Counties with less than 100,000 population

The smaller counties do not show an improvement with the adjusted census. (Table 2) Large counties show a significant improvement, particularly when a more realistic assumption of correlation is simulated, that is, 50 and 75%. (Table 3)

What is the effect of variation in the error parameters on the loss function analysis for states and congressional districts?

Assumptions:

- Reduction of 10% in processing error compared to 1990
- Correlation bias range - 10% to 75% for Blacks only
- Two ways 10 percent change in data collection error

The loss function results show an improvement from adjustment for congressional districts even for small levels of correlation bias.

State results are more sensitive to the assumption of correlation bias.

Table 1. U. S. Summary of Effect of Correlation Bias on Loss Functions for Counties

Distribution Method	Correlation Bias Model	Total Census Population	Total Actual ACE Population	Total Sim ACE Population	Total Target Population		Weighted Levels	Equal CD Shares
GRODSE	10% Correlation Bias Black Only, 90% Processing Error, Revised DA	281,421,906	284,683,795	284,678,082	282,900,984	% Difference	-25.21%	22.81%
						Census Loss / ACE Loss	0.75	1 23
	20% Correlation Bias Black Only, 90% Processing Error, Revised DA	281,421,906	284,683,795	284,678,082	282,975,107	% Difference	-14.69%	36.10%
						Census Loss / ACE Loss	0.85	1.36
	50% Correlation Bias Black Only, 90% Processing Error, Revised DA	281,421,906	284,683,795	284,678,082	283,198,112	% Difference	24.30%	70.30%
						Census Loss / ACE Loss	1.24	1.70
	75% Correlation Bias Black Only, 90% Processing Error, Revised DA	281,421,906	284,683,795	284,678,082	283,383,743	% Difference	64.29%	88.05%
						Census Loss / ACE Loss	1.64	1 88
GROSUC	10% Correlation Bias Black Only, 90% Processing Error, Revised DA	281,421,906	284,683,795	284,678,082	282,901,194	% Difference	-34.75%	22.50%
						Census Loss / ACE Loss	0.65	1.22
	20% Correlation Bias Black Only, 90% Processing Error, Revised DA	281,421,906	284,683,795	284,678,082	282,975,619	% Difference	-24.94%	34.53%
						Census Loss / ACE Loss	0.75	1.35
	50% Correlation Bias Black Only, 90% Processing Error, Revised DA	281,421,906	284,683,795	284,678,082	283,199,458	% Difference	11.34%	65.61%
						Census Loss / ACE Loss	1.11	1.66
	75% Correlation Bias Black Only, 90% Processing Error, Revised DA	281,421,906	284,683,795	284,678,082	283,385,773	% Difference	48.38%	82.22%
						Census Loss / ACE Loss	1.48	1.82

Table 2. U. S. Summary of Effect of Correlation Bias on Loss Functions for Counties with Population Less than 100,000

Distribution Method	Correlation Bias Model	Total Census Population	Total Actual ACE Population	Total Sim ACE Population	Total Target Population		Weighted Levels
GRODSE	10% Correlation Bias Black Only, Revised DA	69,489,081	70,186,846	70,185,741	69,506,988	% Difference	-84.46%
						Census Loss / ACE Loss	0.16
	20% Correlation Bias Black Only, Revised DA	69,489,081	70,186,846	70,185,741	69,519,363	% Difference	-83 77%
						Census Loss / ACE Loss	0.16
	50% Correlation Bias Black Only, Revised DA	69,489,081	70,186,846	70,185,741	69,556,536	% Difference	-80.41%
						Census Loss / ACE Loss	0.20
	75% Correlation Bias Black Only, Revised DA	69,489,081	70,186,846	70,185,741	69,587,514	% Difference	-76.01%
						Census Loss / ACE Loss	0 24
GROSUC	10% Correlation Bias Black Only, Revised DA	69,489,081	70,186,846	70,185,741	69,522,330	% Difference	-94.94%
						Census Loss / ACE Loss	0.05
	20% Correlation Bias Black Only, Revised DA	69,489,081	70,186,846	70,185,741	69,534,729	% Difference	-94.38%
						Census Loss / ACE Loss	0.06
	50% Correlation Bias Black Only, Revised DA	69,489,081	70,186,846	70,185,741	69,572,053	% Difference	-91.36%
						Census Loss / ACE Loss	0 09
	75% Correlation Bias Black Only, Revised DA	69,489,081	70,186,846	70,185,741	69,603,142	% Difference	-87.17%
						Census Loss / ACE Loss	0.13

Table 3. U. S. Summary of Effect of Correlation Bias on Loss Functions for Counties with Population Greater than 100,000

Distribution Method	Correlation Bias Model	Total Census Population	Total Actual ACE Population	Total Sim ACE Population	Total Target Population		Weighted Levels
GRODSE	10% Correlation Bias Black Only, Revised DA	211,932,825	214,496,949	214,492,341	213,264,482	% Difference	0.54%
						Census Loss / ACE Loss	1.01
	20% Correlation Bias Black Only, Revised DA	211,932,825	214,496,949	214,492,341	213,326,249	% Difference	17.65%
						Census Loss / ACE Loss	1 18
	50% Correlation Bias Black Only, Revised DA	211,932,825	214,496,949	214,492,341	213,512,062	% Difference	83 56%
						Census Loss / ACE Loss	1.84
	75% Correlation Bias Black Only, Revised DA	211,932,825	214,496,949	214,492,341	213,666,708	% Difference	153.44%
						Census Loss / ACE Loss	2.53
GROSUC	10% Correlation Bias Black Only, Revised DA	211,932,825	214,496,949	214,492,341	213,249,490	% Difference	-8.76%
						Census Loss / ACE Loss	0.91
	20% Correlation Bias Black Only, Revised DA	211,932,825	214,496,949	214,492,341	213,311,433	% Difference	6.95%
						Census Loss / ACE Loss	1.07
	50% Correlation Bias Black Only, Revised DA	211,932,825	214,496,949	214,492,341	213,497,863	% Difference	66.92%
						Census Loss / ACE Loss	1.67
	75% Correlation Bias Black Only, Revised DA	211,932,825	214,496,949	214,492,341	213,653,039	% Difference	129.71%
						Census Loss / ACE Loss	2.30

Table 4. U. S. Summary of Effect of Data Collection Error by Correlation Bias on Loss Functions for States and Congressional Districts

Geography	Distribution Method	Correlation Bias Model	Total Census Population	Total Actual ACE Population	Total Sim ACE Population	Total Target Population		Weighted Levels	Weighted Shares	Equal CD Shares
State	GRODSE	10% Correlation Bias Black Only, 90% Processing Error, 90% Data Collection Error, Revised DA	281,421,906	284,683,794	284,678,078	282,963,840	% Difference	-11.57%	84 44%	N/A
							Census Loss / ACE Loss	0.88	1 84	N/A
		20% Correlation Bias Black Only, 90% Processing Error, 90% Data Collection Error, Revised DA	281,421,906	284,683,794	284,678,078	283,038,029	% Difference	2 94%	87 42%	N/A
							Census Loss / ACE Loss	1.03	1.87	N/A
		50% Correlation Bias Black Only, 90% Processing Error, 90% Data Collection Error, Revised DA	281,421,906	284,683,794	284,678,078	283,261,300	% Difference	61.94%	91 03%	N/A
							Census Loss / ACE Loss	1 62	1.91	N/A
		75% Correlation Bias Black Only, 90% Processing Error, 90% Data Collection Error, Revised DA	281,421,906	284,683,794	284,678,078	283,447,269	% Difference	133 26%	88.25%	N/A
							Census Loss / ACE Loss	2.33	1.88	N/A
Congressional District	GRODSE	10% Correlation Bias Black Only, 90% Processing Error, 90% Data Collection Error, Revised DA	281,421,906	284,683,787	284,678,060	282,963,842	% Difference	-6 40%	-8.47%	16.90%
							Census Loss / ACE Loss	0 94	0 92	1.17
		20% Correlation Bias Black Only, 90% Processing Error, 90% Data Collection Error, Revised DA	281,421,906	284,683,787	284,678,060	283,038,025	% Difference	8 28%	1.72%	28.93%
							Census Loss / ACE Loss	1.08	1.02	1.29
		50% Correlation Bias Black Only, 90% Processing Error, 90% Data Collection Error, Revised DA	281,421,906	284,683,787	284,678,060	283,261,305	% Difference	62.87%	28 12%	58.02%
							Census Loss / ACE Loss	1.63	1.28	1.58
		75% Correlation Bias Black Only, 90% Processing Error, 90% Data Collection Error, Revised DA	281,421,906	284,683,787	284,678,060	283,447,267	% Difference	117.13%	41.99%	70.82%
							Census Loss / ACE Loss	2.17	1.42	1.71
State	GROSUC	10% Correlation Bias Black Only, 90% Processing Error, 90% Data Collection Error, Revised DA	281,421,906	284,683,794	284,678,078	282,964,036	% Difference	-19 80%	17.69%	N/A
							Census Loss / ACE Loss	0 80	1.18	N/A
		20% Correlation Bias Black Only, 90% Processing Error, 90% Data Collection Error, Revised DA	281,421,906	284,683,794	284,678,078	283,038,482	% Difference	-6.10%	19.99%	N/A
							Census Loss / ACE Loss	0.94	1 20	N/A
		50% Correlation Bias Black Only, 90% Processing Error, 90% Data Collection Error, Revised DA	281,421,906	284,683,794	284,678,078	283,262,480	% Difference	49 24%	25 45%	N/A
							Census Loss / ACE Loss	1 49	1.25	N/A
		75% Correlation Bias Black Only, 90% Processing Error, 90% Data Collection Error, Revised DA	281,421,906	284,683,794	284,678,078	283,448,937	% Difference	115 29%	27.98%	N/A
							Census Loss / ACE Loss	2.15	1.28	N/A
Congressional District	GROSUC	10% Correlation Bias Black Only, 90% Processing Error, 90% Data Collection Error, Revised DA	281,421,906	284,683,787	284,678,060	282,964,032	% Difference	-15 35%	-32.92%	4.26%
							Census Loss / ACE Loss	0.85	0.67	1.04
		20% Correlation Bias Black Only, 90% Processing Error, 90% Data Collection Error, Revised DA	281,421,906	284,683,787	284,678,060	283,038,475	% Difference	-1 50%	-23 32%	15.81%
							Census Loss / ACE Loss	0.99	0.77	1.16
		50% Correlation Bias Black Only, 90% Processing Error, 90% Data Collection Error, Revised DA	281,421,906	284,683,787	284,678,060	283,262,484	% Difference	50.09%	3 85%	45.16%
							Census Loss / ACE Loss	1.50	1 04	1.45
		75% Correlation Bias Black Only, 90% Processing Error, 90% Data Collection Error, Revised DA	281,421,906	284,683,787	284,678,060	283,448,933	% Difference	101.36%	20 55%	59.62%
							Census Loss / ACE Loss	2 01	1.21	1.60

Geography	Distribution Method	Correlation Bias Model	Total Census Population	Total Actual ACE Population	Total Sim ACE Population	Total Target Population		Weighted Levels	Weighted Shares	Equal CD Shares
State	GRODSE	10% Correlation Bias Black Only, 90% Processing Error, 110% Data Collection Error, Revised DA	281,421,906	284,683,794	284,678,078	282,838,206	% Difference	-33 12%	67.69%	N/A
							Census Loss / ACE Loss	0.67	1.68	N/A
		20% Correlation Bias Black Only, 90% Processing Error, 110% Data Collection Error, Revised DA	281,421,906	284,683,794	284,678,078	282,912,239	% Difference	-21.96%	70.91%	N/A
							Census Loss / ACE Loss	0.78	1.71	N/A
		50% Correlation Bias Black Only, 90% Processing Error, 110% Data Collection Error, Revised DA	281,421,906	284,683,794	284,678,078	283,135,062	% Difference	23.71%	76 30%	N/A
							Census Loss / ACE Loss	1.24	1.76	N/A
		75% Correlation Bias Black Only, 90% Processing Error, 110% Data Collection Error, Revised DA	281,421,906	284,683,794	284,678,078	283,320,603	% Difference	79 63%	75.74%	N/A
							Census Loss / ACE Loss	1.80	1 76	N/A
Congressional District	GRODSE	10% Correlation Bias Black Only, 90% Processing Error, 110% Data Collection Error, Revised DA	281,421,906	284,683,787	284,678,060	282,838,205	% Difference	-25.73%	-10 22%	12.55%
							Census Loss / ACE Loss	0.74	0 90	1.13
		20% Correlation Bias Black Only, 90% Processing Error, 110% Data Collection Error, Revised DA	281,421,906	284,683,787	284,678,060	282,912,237	% Difference	-13 83%	-0.56%	24.04%
							Census Loss / ACE Loss	0.86	0 99	1.24
		50% Correlation Bias Black Only, 90% Processing Error, 110% Data Collection Error, Revised DA	281,421,906	284,683,787	284,678,060	283,135,064	% Difference	31.43%	25 24%	52.81%
							Census Loss / ACE Loss	1 31	1.25	1.53
		75% Correlation Bias Black Only, 90% Processing Error, 110% Data Collection Error, Revised DA	281,421,906	284,683,787	284,678,060	283,320,598	% Difference	78 30%	39.47%	66.43%
							Census Loss / ACE Loss	1.78	1.39	1.66
State	GROSUC	10% Correlation Bias Black Only, 90% Processing Error, 110% Data Collection Error, Revised DA	281,421,906	284,683,794	284,678,078	282,838,442	% Difference	-40.57%	-0.99%	N/A
							Census Loss / ACE Loss	0.59	0.99	N/A
		20% Correlation Bias Black Only, 90% Processing Error, 110% Data Collection Error, Revised DA	281,421,906	284,683,794	284,678,078	282,912,780	% Difference	-30.08%	1.31%	N/A
							Census Loss / ACE Loss	0.70	1 01	N/A
		50% Correlation Bias Black Only, 90% Processing Error, 110% Data Collection Error, Revised DA	281,421,906	284,683,794	284,678,078	283,136,420	% Difference	12.57%	7.73%	N/A
							Census Loss / ACE Loss	1.13	1 08	N/A
		75% Correlation Bias Black Only, 90% Processing Error, 110% Data Collection Error, Revised DA	281,421,906	284,683,794	284,678,078	283,322,628	% Difference	64.09%	11.92%	N/A
							Census Loss / ACE Loss	1.64	1.12	N/A
Congressional District	GROSUC	10% Correlation Bias Black Only, 90% Processing Error, 110% Data Collection Error, Revised DA	281,421,906	284,683,787	284,678,060	282,838,447	% Difference	-33 85%	-34.57%	0.05%
							Census Loss / ACE Loss	0 66	0 65	1.00
		20% Correlation Bias Black Only, 90% Processing Error, 110% Data Collection Error, Revised DA	281,421,906	284,683,787	284,678,060	282,912,781	% Difference	-22.66%	-25 68%	10 89%
							Census Loss / ACE Loss	0.77	0.74	1 11
		50% Correlation Bias Black Only, 90% Processing Error, 110% Data Collection Error, Revised DA	281,421,906	284,683,787	284,678,060	283,136,422	% Difference	19 82%	0.12%	39.25%
							Census Loss / ACE Loss	1 20	1.00	1.39
		75% Correlation Bias Black Only, 90% Processing Error, 110% Data Collection Error, Revised DA	281,421,906	284,683,787	284,678,060	283,322,632	% Difference	63 71%	16 66%	54.14%
							Census Loss / ACE Loss	1.64	1.17	1.54

ESCAP MEETING NO. 46 - 02/26/01

MINUTES

Minutes of the Executive Steering Committee on Accuracy and Coverage Evaluation (A.C.E.) Policy (ESCAP) Meeting # 46

February 26, 2001

Prepared by: Sarah Brady

The forty-sixth meeting of the Executive Steering Committee on Accuracy and Coverage Evaluation Policy was held on February 26, 2001 at 10:30 and 1:30. The agenda for the 10:30 meeting was to discuss remaining demographic analysis issues. The agenda for the 1:30 meeting was to present loss function results for counties and to examine the impact of synthetic bias on the loss functions.

Committee Attendees:

Nancy Potok
Paula Schneider
Cynthia Clark
Nancy Gordon
John Thompson
Jay Waite
Bob Fay
Howard Hogan
Ruth Ann Killion
John Long
Carol Van Horn

Deputy Director/Acting Director:
William Barron

Other Attendees:

Marvin Raines
Bill Bell
Donna Kostanich (PM only)
Alfredo Navarro (PM only)
Richard Griffin (PM only)
Donald Malec (PM only)
Maria Urrutia
Sarah Brady

I. Demographic Analysis

John Long passed out sections of the DSSD Census 2000 Procedures and Operations Memorandum Series B-4: Accuracy and Coverage Evaluation: Demographic Analysis Results. John described how the based demographic analysis (DA) and alternative DA will be presented in the document. The handouts are attached.

II. Distribution of the Difference between Adjusted and Unadjusted

John Long presented data on the distribution of the numeric difference between the adjusted and the unadjusted census for places, counties, county subdivisions, and American Indian Reservations based upon the population of the entity. The Committee expressed an interest to see the data presented for the same entities but in terms of a percent difference in population. John will provide these data to the Committee at a later time.

III. Loss Functions for Counties

Alfredo Navarro presented loss function results for counties. The data were part of the sensitivity analysis at the county level for loss functions. For the purpose of this analysis, the total error model assumptions were: correlation bias ranging from 10 percent to 75 percent for the Black population only, processing error was reduced by 10 percent as compared to 1990, and data collection error was examined at a 10 percent reduction and a 10 percent increase as compared to 1990.

Improvements in accuracy were noted for both numeric and distributive accuracy for the universe consisting of all counties, which are consistent with the findings for states and congressional districts. However, for counties with a population of less than 100,000, the loss functions indicated that the adjusted was less accurate than the unadjusted census.

IV. Synthetic Error

Richard Griffin presented data illustrating the effect of synthetic error on loss function analysis.

- The question this analysis addressed was the effect on loss functions from synthetic error.
- Both the census and the A.C.E. are subject to synthetic error.
- Synthetic error is added to the loss functions to determine if the increase in loss is disproportionate, therefore, favoring the census or adjustment.

We compared the effect of adding synthetic error on the census and adjusted losses. The Committee concluded that for the loss functions with which we are concerned, the weighted squared error and the equal congressional district loss functions, the largest effect observed favoring the census was a 58 percent increase for the weighted squared error loss for state shares. Therefore, the Committee will take this into account when examining the loss function results.